Journal of Applied Psychology

JOHN G. DARLEY, Editor
UNIVERSITY OF MINNESOTA

LORRAINE BOUTHILET, Managing Editor

Table of Contents

The Effects of Real-Life Motivation on Questionnaire Response: A. Heron 65

Permanence of Measured Vocational Interests of Adult Males: M. K. Powers 69

Influence of Scale Administrator on Employee Attitude Responses: M. D. Dunnette and H. G. Heneman, Jr. 73

Attitudes of Union and Non-Union Employees in a Calcutta Electrical Engineering Factory: H. C. Ganguli 78

Prediction of Success from Examination of Performance During the Training Period: P. C. Smith and R. A. Gold 83

The Measurement of Personal Factors Related to Success of Office Workers: M. R. Guthrie .. 87

Internal Relations of Elemental Motions Within a Task: N. B. Hall, Jr. 91

Factorial Analysis of Complex Psychomotor Performance and Related Skills: E. A. Fleishman and W. E. Hempel, Jr. 96

Spatial Factors in Check Reading of Dial Groups: R. S. Lincoln and E. Averbach 105

The Contribution of Lecture Supplements to the Effectiveness of an Attitudinal Film: F. T. Staudohar and R. G. Smith, Jr. 109

Leadership and Predictive Abstracting: C. G. Browne and R. P. Shore 112

Validity of Extrapolating Nonresponse Bias from Mail Questionnaire Follow-ups: H. Zimmer. 117

A Biographical Inventory for Students: II. Validation of the Instrument: L. Siegel 122

Equivalence of Forms of the Wonderlic Personnel Test: A Study of Reliability and Interchangeability: H. B. Weaver and C. A. Boneau 127

Experimental Manipulation of the Halo Effect: D. M. Johnson and R. N. Vidulich 130

Comparison of Two Visual Display Presentations: A. V. Churchill 135

American Psychological Association

Volume 40, Number 2 April, 1956

Consulting Editors

Harold E. Burtt, *Ohio State University*
Alphonse Chapanis, *Johns Hopkins University*
Clifford E. Jurgensen, *Minneapolis Gas Company*
Laurence S. McGaughran, *University of Houston*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
Harold F. Rothe, *Fairbanks, Morse and Company*
Julian B. Rotter, *Ohio State University*
Thomas A. Ryan, *Cornell University*
Donald E. Super, *Columbia University*
Miles A. Tinker, *University of Minnesota*
Alfred C. Welch, *University of New Mexico*

This journal gives primary consideration to original investigations in any field of applied psychology except clinical and consulting psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: vocational and educational prognosis, diagnosis, and guidance at the secondary and college level; personnel research in business, industry, and government; biomechanics; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research.

Because of the large number of manuscripts submitted, authors should adhere to the rule of "brevity consistent with clarity." The typical manuscript should run to approximately 4,000 words. There is a lag of approximately twelve months between receipt and publication of an article. Authors may request advanced publication if they are prepared to pay the cost of printing the necessary extra pages.

Manuscripts should be addressed to the Editor, John G. Darley, 408 Johnston Hall, University of Minnesota, Minneapolis 14, Minnesota. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies.

Manuscripts must conform to the style requirements described in the "Publication Manual of the American Psychological Association," *Psychol. Bull.*, 1952, **49**, No. 4, Part 2.

Journal of Applied Psychology

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W.
Washington 6, D. C.

$8.00 per volume $1.50 per issue

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the subscription office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947.

Copyright © 1956 by the American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 40, No. 2 APRIL, 1956

The Effects of Real-Life Motivation on Questionnaire Response [1]

Alastair Heron

Medical Research Council, London

There is abundant evidence in this Journal (2, 4, 8) and elsewhere (1, 5, 9, 10, 13, 14) that if university and college students are instructed to do so, they can improve their score on most adjustment questionnaires and inventories. As far as the present writer can find, there is remarkably little evidence concerning what actually happens in real life. Green (6) published a valuable study in 1951, quoting two other papers, one concerned like himself with a police situation (12), the other based on a single case history (11). Green's study showed that in a police selection situation an applicant group produced significantly more favorable mean scores than did a "secure tenure" group on sections S, T, D, and C of Guilford's STDCR Inventory, on Section A of his GAMIN Inventory, and on the Mechanical, Scientific, Persuasive, Artistic, and Social Service sections of the Kuder Preference Record. In presenting his material, however, Green made it clear that he found it difficult to ensure that his two groups were in every respect comparable. His samples were also rather small (N's = 45 and 70). The present paper presents the results of a fairly large-scale experiment designed from the outset to test the hypothesis that with personality-questionnaire material, response distributions under selection conditions will differ significantly from those obtained under research conditions, all other variables having been fully controlled.

[1] Paper presented to the Twelfth International Congress of Applied Psychology, London, July, 1955.

Method

Subjects. The *Ss* for this experiment were 400 male applicants for the job of omnibus conductor, between the ages of 18 and 45 and without previous experience in this type of work. From test data available on a comparable sample, it could be inferred that intelligence ranged from well above average for the general population, to literate subjects just above the borderline of statutory mental defect. Such applicants would come from the same socio-economic background as the great bulk of unskilled and semi-skilled factory workers.

The questionnaire material. The instrument employed was a two-part personality measure designed by the writer (7) for use as a research criterion suited to the varied needs of psychologists in different fields. The first part consists of 74 items in statement form, of which 20 items form the basis for a measure of emotional maladjustment. The majority of these 20 items, selected after wide use in the field, repeated internal and external validation, and thorough item analysis, were drawn from psychosomatic inventories such as the MMPI and the Maudsley Medical Questionnaire (3). Apart from eight items of the MMPI Lie Scale, the remaining items of this first part are buffers, concerned with physical health and intended to convey the general appearance of a health inventory. The second part, given after an interval during which some other activity is inserted, consists of 36 items, 12 of which form the basis for a measure of sociability. The 12 selected items, like those of the remaining 24, were drawn from Guilford's R (rhathymia) scale. In addition to other criteria for selection, it was provided that in order to contribute to the total score for the second part, an item must not achieve a significantly nonchance distribution in respect of emotional maladjustment, and also that the items selected must not, when considered collectively, achieve a significantly nonchance additive chi square in respect of their distribution for emotional maladjustment. The object was of course to ensure maximum independence of each other for the two measures concerned.

The split-half self-correlations, corrected to full length, were .81 for emotional maladjustment (20 items) and .74 for sociability (12 items), with $N = 378$. With a cutting point placed one sigma above the mean for 251 normal respondents and one sigma below the mean for 27 hospitalized civilian neurotics of mixed diagnoses, all responding under research conditions, misclassification was 13 per cent for the measure of emotional maladjustment.

This two-part inventory was prepared so that it could be administered in two ways. One method of administration was the conventional pencil-and-paper form, with "true" and "not true" alternatives; the other made use of a box-and-card technique. Each item was printed on a card identical with those used in the card presentation of the MMPI, but all cards were then sealed in the central section of a three-part wooden container. The *S* was shown how to withdraw one card at a time from a slot at the base of this section, and to post it in either the "true" or "not true" box on one side or the other of the card-storage section. He did not know how many cards he would have to do, nor could he consult earlier or later cards before responding to the one in his hand, and he could not change his mind about an earlier response. In both methods of administration, general instructions were identical, and in particular it was emphasized that if difficulty was encountered in making up one's mind about a particular item, the response should be chosen which was "more nearly or most often the case, even if only by ever so little."

Design of the experiment. The selection situation in the recruiting center of the public transport department concerned lent itself admirably to a simple design. All applicants (for a variety of jobs) were seen first for a few minutes by one or other of two interviewers in the personnel department. Those considered likely to justify further expenditure of time by themselves and the department were then asked to fill out application forms under group supervision. Applicants then went before one or other of three interviewers and their job desires were discussed fully. No shortage of vacancies for the job of omnibus conductor prevailed at any time during the course of this study. If a man was considered suitable for this job, he was told that if he passed the medical examination and the vision tests, he would be accepted for the training school. He then passed to the Medical Department waiting room, whence he was soon called forward for examination. Those passing the medical examination then moved forward to the Allocation Room, where they were told they had got the job, and were given a card with instructions for reporting to the Training School. It will be seen that in the eyes of a job applicant there were three hurdles: the initial brief interview, the application form and its consideration by the second interviewer, and the examination by the Medical Department. When—and only when—he reached the Allocation Room did he know he had got the job and had no more hurdles.

Two hundred men were given the two-part personality measure as part of the Medical Department procedure by a white-coated orderly or by the writer acting as one. Two hundred more men were (by alternation) allowed to pass through the Medical Department *without* being given the two-part measure. (Fully adequate precautions were taken to ensure that these men did not see or know about the activities with the measure on the part of their fellow applicants. This was facilitated, of course, by the fact that applicants for a variety of jobs were passing through simultaneously.) These two hundred were not approached until they were leaving the Allocation Room, having obtained the job. They were then asked if they were willing to spare a few more minutes to assist in a piece of medical research being carried out independently of the public transport undertaking, and forming part of a large-scale

Table 1

Means and Variances for All Scores, as Described in Text

	Emotional Maladjustment				Sociability			
	Box		Paper		Box		Paper	
	Mean	σ^2	Mean	σ^2	Mean	σ^2	Mean	σ^2
Research (main)	5.75	9.36	5.56	12.81	4.51	5.06	4.50	4.25
N	100		100		100		100	
Selection (main)	3.61	6.25	3.65	7.56	4.40	4.84	4.60	3.24
N	100		100		100		100	
Selection (check)	3.65	7.40	3.80	11.97	4.43	2.43	4.58	3.13
N	147		145		147		145	

Note.—1. Mean scores for emotional maladjustment differ significantly ($P = .001$) when compared for *conditions* of administration (Research [main] vs. either Selection [main] or Selection [check]). No other means differ significantly.
2. The only variances to differ significantly ($P = .05$) are 7.40 vs. 11.97 (Selection [check], box vs. paper on Emotional Maladjustment) and 4.84 vs. 3.24 (Selection [main], box vs. paper, on Sociability).

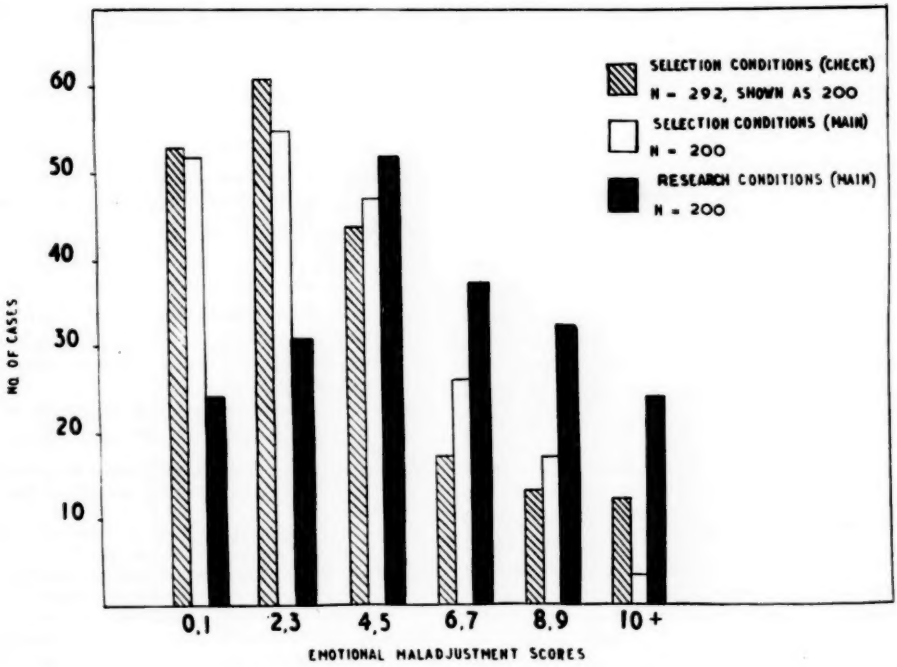

FIG. 1. Score distributions for emotional maladjustment.

enquiry into matters concerned with industrial health. (This explanation was completely genuine.) Only 204 men had to be asked in order to obtain the required 200.

Of each group of 200 men, 100 were given the two-part measure by the conventional pencil-and-paper method, and the other 100 by the new box-and-card method. This design accepted the risk that there might prove to be an interaction between method of administration and circumstances of response. It seemed essential, however, to use this opportunity for comparing the two methods of administration, and it was, of course, evident that if the results did *not* differ for administration but only for circumstances, the main point of the experiment would have been more powerfully made.

It was not possible to replicate the entire experiment elsewhere, but some valuable data were obtained from the personnel department of a public transport undertaking in another part of the country. This information concerned responses obtained under routine selection conditions in respect of the same type of work as in the main study. Applicants were given the two-part measure at an earlier stage in the procedure, but otherwise comparably. Vacancies were available as generally as in the main study. One hundred forty-five men completed the measure by the pencil-and-paper method, and 147 by box-and-cards. These data can therefore be appended as an external check on those shown under "selection" conditions in the main study.

Results and Conclusions

The means and variances for all scores are shown in Table 1, and score distributions for emotional maladjustment appear in Fig. 1.

It can be seen by inspection, without recourse to formal tests of significance, that:

1. The 12-item score for sociability is apparently unaffected by method of administration or by circumstances of response;

2. The 20-item score for emotional maladjustment is unaffected by method of administration but is seriously affected by circumstances of response, the means and variances under selection circumstances being markedly and consistently lower than those prevailing under research circumstances.

From Fig. 1 [2] it will be clear that there is no tendency for high scores only to be affected, with a resultant median bunching, but rather for a general downward shift throughout the scoring range. It is of considerable interest to note that under research conditions 24 men out of 200 (12 per cent of the sample) obtain scores of 10 or greater; with the cutting point described earlier, these would be regarded as "probably maladjusted." This figure of 12 per cent accords well with other data obtained in earlier work carried out by the Unit to which the writer belongs, and by others. By contrast, under selection conditions only three men out of 200 (1½ per cent of the main sample) appear as "probably maladjusted," a patently unlikely figure in such a sample.

The hypothesis stated above is therefore sustained by the experiment, and would appear to justify the conclusion that adjustment inventories and psychosomatic questionnaires are responded to differently when in a real-life situation there is some incentive not to admit personality defects. This finding accords with those of Green (**6**), as does the negative finding concerning the sociability score, this being based entirely on items drawn from Guilford's R scale, which alone of the STDCR sections showed no significant differences in Green's study.

Summary

Four hundred men between the ages of 18 and 45 applying for the job of omnibus conductor were given a two-part personality measure, one part concerned with emotional maladjustment, the other with sociability. An experiment was designed to test the hypothesis that response distributions under selection conditions would differ significantly from those obtained under research conditions, when all other variables were fully controlled. The hypothesis was clearly sustained for emotional maladjustment and refuted for sociability. Subsidiary and fully comparable data partially replicated the experiment and served to confirm it. The results are in accord with the only other comparable study known to the writer, and appear to provide long-awaited evidence, concerning the consequences of real-life motivation on questionnaire response, of a kind that could not safely be inferred from many previous artificial studies of "faking."

Received May 11, 1955.

[2] To permit direct comparison with the two main samples, figures relating to the check sample have been reduced proportionately from a basis of $N = 292$ to one of $N = 200$.

References

1. Cross, O. H. A study of faking on the Kuder Preference Record. *Educ. psychol. Measmt,* 1950, **10**, 271–277.
2. Ellis, A. Recent research with personality inventories. *J. consult. Psychol.,* 1953, **17**, 45–49.
3. Eysenck, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
4. Faw, V. Situational variations of neurotic scores. *J. consult. Psychol.,* 1948, **12**, 255–258.
5. Gough, H. G. Simulated patterns on the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.,* 1947, **42**, 215–225.
6. Green, R. F. Does a selection situation induce testees to bias their answers on interest and temperament tests? *Educ. psychol. Measmt,* 1951, **11**, 503–515.
7. Heron, A. A two-part personality measure for use as a research criterion. *Brit. J. Psychol.,* 1956, **47**, in press.
8. Hunt, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *J. consult. Psychol.,* 1948, **12**, 396–402.
9. Kelly, E. L., Miles, C. C., & Terman, L. M. Ability to influence one's score on a pencil and paper test of personality. *Charact. & Pers.,* 1936, **4**, 206–215.
10. Mais, R. D. Fakability of the classification inventory scored for self confidence. *J. appl. Psychol.,* 1951, **35**, 172–174.
11. Paterson, D. G. Vocational interest inventories in selection. *Occupations,* 1946, **25**, 152–153.
12. Searless, J. R., & Leonard, J. M. *Experiments in the mental testing of Detroit policemen.* Detroit: Detroit Bureau of Government Research, 1936.
13. Steinmetz, H. L. Measuring ability to fake occupational interest. *J. appl. Psychol.,* 1932, **16**, 123–130.
14. Wesman, A. G. Faking personality test scores in a simulated employment situation. *J. appl. Psychol.,* 1952, **36**, 112–113.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Permanence of Measured Vocational Interests of Adult Males [1]

Mabel K. Powers

University of Minnesota

All previous published studies on permanence of vocational interests as measured by the Strong Vocational Interest Blank (SVIB) have shown remarkable permanence of interests (**3**, **4**, **5**). However, in all these studies the subjects have been high school or college students at the time of the first administration of the SVIB. It was hypothesized that results on adult males representing all levels of occupations and a wide range in age might be quite different, especially if the time between testings was a period of unemployment and economic disruption. The present study was designed to test this hypothesis.

Method

Subjects. The subjects (*Ss*) of this study were 109 men who were in the sample studied by the Minnesota Employment Stabilization Research Institute in 1931 (**2**) and again in 1941 as part of a follow-up study of the same Institute. In 1931 the 109 men ranged in age from 16 to 63 years with a mean age of 33.7 years. Sixty-three of the 109 were in the skilled, semiskilled, and unskilled occupational levels, 16 were in lower-level clerical occupations, and the remaining 30 were in professional, managerial, high-level business, and sales occupations. Ninety-four were unemployed in 1931 and all of the men in the sample were unemployed at some time during the 10 years between testings.

Procedure. Since the 1927 form of the SVIB was used in the 1931 testing, the answers on the original blanks were transferred to Hankes answer sheets after eliminating the 20 items not appearing on the 1938 revision of the test. Both tests for each individual were scored on 44 occupational scales. A rank-difference correlation coefficient was computed for each *S* using the rank order of the standard scale scores on the 44 occupational scales at each testing. In addition, a product-moment correlation coefficient was computed between total standard scores for the 109 *Ss* on each of the 44 occupational scales for the first and second administration of the SVIB. Also, the differences between means of the 109 *Ss* on each scale for the two testings were computed and checked for significance of differences. For those scales where the difference between mean standard scores was significant, percentage of overlap was computed.

All answer sheets were then analyzed for the presence of Primary (P), Secondary (S), Tertiary (T), and Reject (R) patterns in each of the 11 interest groups, using Darley's (**1**) method for the P, S, and T patterns and classifying all others as Rejects. A tabulation was made of the difference in patterns on the first and second testings for each *S* in each of the 11 interest groups, assigning a difference of 0, 1, 2, or 3 depending on the change in pattern between the two testings. To illustrate, if an *S* had a Primary pattern in Group I in the first testing and a Secondary pattern on the second testing, he was assigned a difference of one for that group; if he had a Primary pattern in Group I in the first testing and a Reject pattern in the second testing, he was assigned a difference of three. These differences, regardless of sign, were summed for each individual and the resulting total was considered the over-all difference (D) in profile for each subject. This simple test could not be checked for significance because the distribution is not known for the measure of similarity used. Therefore, the mean of the differences for the *Ss* in this sample was compared with the greatest possible difference for any two SVIB profiles and also with the greatest possible average difference for the 109 *Ss* in this sample based on actual differences.

The *Ss* were assigned to subsamples on the basis of D, using $+1$ *SD*, mean, and -1 *SD* as cutting points; the *Ss* were also assigned to subsamples on the basis of the rank-order correlation coefficients based on 44 occupational scales, using the same cutting points.

Results

Test-retest correlations for individual profiles. When rho coefficients of correlation

Table 1

Distribution of Rho Coefficients of Correlation Between Standard Scores

rho	*N*	rho	*N*
.80–.99**	56	.00–.19	2
.60–.79**	37	−.20–−.09	1
.40–.59**	7	−.40–−.19*	2
.20–.39***	4		

* One rho significantly different from 0.0 at .05 level.
** Significantly different from 0.0 at the .01 level.
*** One rho significantly different from 0.0 at the .05 level and one at the .01 level.
Note.—Median rho = .80, mean rho = .78; mean rho + 1 *SD* = .90, mean rho − 1 *SD* = .56.

[1] This paper is based upon a portion of the author's doctoral dissertation which was under the direction of Prof. D. G. Paterson.

Table 2

Distribution of Correlation Coefficients Between Test-Retest Standard Scores

Occupational Scale	r*	Occupational Scale	r*
Artist	.74	Personnel director	.62
Psychologist	.73	Public administrator	.57
Architect	.76	Y.M.C.A. secretary	.61
Physician	.76	Social science high school teacher	.67
Osteopath	.56	City school superintendent	.59
Dentist	.69	Minister	.59
Veterinarian	.70	Musician	.65
Mathematician	.75	C.P.A.	.74
Physicist	.81	Senior C.P.A.	.57
Engineer	.82	Accountant	.78
Chemist	.81	Office man	.74
Production manager	.69	Purchasing agent	.67
Farmer	.75	Banker	.82
Aviator	.72	Mortician	.65
Carpenter	.81	Pharmacist	.47
Printer	.65	Sales manager	.62
Mathematics-physical science teacher	.61	Real estate salesman	.44
Industrial arts teacher	.72	Life insurance salesman	.66
Vocational agriculture teacher	.62	Advertising man	.73
Policeman	.72	Lawyer	.77
Forest service man	.62	Author-journalist	.74
Y.M.C.A. physical director	.62	President manufacturing concern	.64

* All r's significantly different from zero at .001 level.

were computed between ranks for each of the 44 occupational scales in each testing for each individual, the median rho was + .80, the mean rho was + .78 with an *SD* of 12. The range of rhos was from + .96 to − .38 with the distribution markedly negatively skewed. For 101 of the 109 *S*s, rho was significantly different (beyond 1%) from zero; for 18, or 16.5 per cent, rho was + .90 or above. The distribution of rhos is shown in Table 1.

Test-retest correlations for total sample. Product-moment correlation coefficients between total test-retest scores on individual scales ranged from + .82 on the Engineer scale to + .44 on the Real Estate Salesman scale with all r's being significantly different from zero (at .001 level). By means of transformation to Z coefficients, the average r for all 44 scales was found to be + .69. The correlations are given in Table 2.

Difference in mean scores. The range of differences between mean standard scores on separate scales was from − .2 standard scores on Purchasing Agent to + 4.5 standard scores on Aviator. The differences were significant at the .01 level for 14 scales and at the .05 level for six additional scales. When percentage of overlap was computed for the distributions of scores on these 20 scales, using Tilton's (6) Method, the smallest percentage of overlap found for any scale was 87 for the Osteopath and Aviator scales. The differences between means and percentage of overlap where differences are significant are shown in Table 3.

Permanence of group patterns. For the 109 *S*s in this sample, there were 182 P, 150 S, 518 T, and 349 R patterns for the 1931 administration and 218 P, 138 S, 515 T, and 348 R patterns for the 1941 administration. The distribution by Groups is shown in Table 4. Since the greatest possible difference for P's and R's is three and for S's and T's is two, the greatest possible difference (D) for the whole sample was 2929. The actual difference was 631, just slightly more than 20% of the possible difference.

The greatest possible difference between two profiles on any single individual is 33 if all original patterns are P's and/or R's; if all

Table 3

Difference in Means for Each Scale on Two Administrations of SVIB ($N = 109$); Per Cent of Overlap where Differences are Significant

Occupational Scale	d	% of Overlap	Occupational Scale	d	% of Overlap
Artist	.9	—	Personnel director	.6	—
Psychologist	2.6**	90	Public administrator	3.1**	89
Architect	1.1	—	Y.M.C.A. secretary	−2.9*	91
Physician	2.3**	93	Social science high school teacher	−2.9**	92
Osteopath	3.6**	87	City school superintendent	−1.4	—
Dentist	.8	—	Minister	− .4	—
Veterinarian	2.3**	98	Musician	− .6	—
Mathematician	.8	—	C.P.A.	−1.4*	95
Physicist	2.7**	91	Senior C.P.A.	.7	—
Engineer	4.0**	87	Accountant	−1.7*	95
Chemist	2.7**	92	Office man	−2.4*	93
Production manager	2.6**	90	Purchasing agent	− .2	—
Farmer	− .4	—	Banker	−1.5*	94
Aviator	4.5**	87	Mortician	.3	—
Carpenter	.8	—	Pharmacist	.9	—
Printer	.3	—	Sales manager	−1.3	—
Mathematics-physical science teacher	− .3	—	Real estate salesman	−1.8*	90
Industrial arts teacher	3.1**	93	Life insurance salesman	−2.6**	89
Vocational agriculture teacher	.5	—	Advertising man	− .3	—
Policeman	.7	—	Lawyer	.7	—
Forest service man	.4	—	Author-journalist	.6	—
Y.M.C.A. physical director	−1.7	—	President manufacturing concern	2.2**	91

* Significant at .05 level.
** Significant at .01 level.

original patterns are S's or T's, the greatest possible difference is 22. Since the greatest possible difference for the 109 *S*s in this study was 2929, the average for any one *S* would have been 27. The actual differences ranged from one to 13 with a mean difference of 5.8, median difference of 6.0. The actual distribution of differences (D) is shown in Table 5.

The chi square computed to test the relationship between permanence of interests as measured by rank-order correlation and by difference in group patterns was 50.82 ($P < .001$).

Discussion

The results of this investigation force a rejection of the hypothesis that adult males, as represented by the sample used, may be expected to exhibit a lack of permanence in vocational interests even if the period between testings is one of widespread unemployment. The remarkably high correlation between two profiles for an individual and between total test-retest standard scores for the group on each of the 44 scales and the high degree of overlap of the distributions of standard scores on each scale and the stability of group patterns all lead to the conclusion that the vocational interests of adult males, regardless of age, aptitudes, education, vocational opportunity, and economic status tend to remain stable for 10 years and presumably over the long span of adulthood. Since the majority of the *S*s were from the low-level clerical, skilled, semiskilled, and unskilled categories, these data contribute information about a large section of the male labor force which is not available from previous studies. Therefore, the results of this study support the conclusion that SVIB may be used with confidence at the lower occupational levels as well as with college and high school students who may be destined for economic competition at the higher occupational levels.

Table 4

Distribution of Primary, Secondary, Tertiary, and Reject Patterns

Group		P	S	T	R
I	1931	3	6	57	43
	1941	7	10	54	38
II	1931	4	7	62	36
	1941	6	6	68	29
III	1931	43	19	35	12
	1941	58	17	28	6
IV	1931	35	16	41	17
	1941	43	23	30	13
V	1931	14	11	51	33
	1941	12	13	52	32
VI	1931	14	11	42	42
	1941	14	16	32	47
VII	1931	6	5	53	45
	1941	5	4	46	54
VIII	1931	25	24	55	5
	1941	25	20	56	8
IX	1931	17	22	51	19
	1941	14	10	56	29
X	1931	5	10	26	68
	1941	10	3	29	67
XI	1931	16	19	45	29
	1941	24	16	44	25
Total	1931	182	150	518	349
	1941	218	138	515	348

Table 5

Distribution of Sums of Differences (D's) Between Patterns

Difference (D)	*N*
13	3
12	3
11	2
10	6
9	5
8	7
7	17
6	13
5	13
4	12
3	13
2	9
1	6
0	0
Total	109

Median = 6.0, mean = 5.8, *SD* = 3.0.

Summary

The *S*s of the study were 109 males from the Minnesota Employment Stabilization Research Institute sample who were tested with the Strong Vocational Interest Test in 1931 and again in 1941. The two tests for each *S* were scored on the 44 occupational keys. The age range of the *S*s was from 16 to 63 years in 1931 and they were generally from the middle and lower socioeconomic groups. Permanence of interests was measured in 4 ways: (*a*) rank-order correlation between test-retest standard scores on the 44 occupational scales for each subject, (*b*) product-moment correlation coefficients between total test-retest standard scores on each of the 44 occupational scales, (*c*) difference between mean test and retest standard score for the total sample for each of the 44 occupational scales and percentage of overlap for those scales showing statistically significant differences, and (*d*) difference in group patterns from test to retest.

The conclusion reached is that vocational interests of adult males representing a wide range in age and socioeconomic status are remarkably stable when permanence is measured by the four methods outlined above.

Received May 13, 1955.

References

1. Darley, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corp., 1941.
2. Paterson, D. G. Research studies in individual diagnosis. *Bull. Employment Stabilization Res. Inst.*, Univer. of Minnesota, 1934, **3**, No. 4.
3. Stordahl, K. E. Permanence of Strong Vocational Interest Blank scores. *J. appl. Psychol.*, 1954, **38**, 423–427.
4. Strong, E. K., Jr. Permanence of interest scores over 22 years. *J. appl. Psychol.*, 1951, **35**, 89–91.
5. Strong, E. K., Jr. Nineteen-year followup of engineer interests. *J. appl. Psychol.*, 1952, **36**, 65–74.
6. Tilton, J. W. The measurement of overlapping. *J. educ. Psychol.*, 1937, **28**, 656–662.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Influence of Scale Administrator on Employee Attitude Responses

Marvin D. Dunnette

Minnesota Mining & Manufacturing Co., St. Paul, Minnesota

and Herbert G. Heneman, Jr.[1]

Industrial Relations Center, University of Minnesota

Systematic investigation of employee attitudes is a relatively recent development in American business and industry. In recent years, employee attitude information has proved useful as a means of upward communications from employees to their employers. Thus employee attitude measurements have been widely used by employers in assessing the relative strengths and/or weaknesses of their personnel and employee relations programs. As an aid in this direction, most employee opinion surveys have adopted content that is very broad in its coverage. For example, standardized employee attitude questionnaires often include specific subscales designed to measure attitudes toward such factors as hours worked, pay received, type of work, working conditions, supervision, coworkers, employee benefits, job security, opportunities for increased status and promotion, adequacy of employment communications, etc. Most research questions concerning the writing of items, the statistical analyses appropriate to item retention, and the final standardization of attitude questionnaires have been answered adequately.

Questions still arise, however, with regard to the optimal methods of scale administration. It is desirable, obviously, that attitude scales be administered in such a way as to secure as accurate a picture as possible of employees' "true" attitudes—their actual unbiased responses to the statements in the particular scale being administered.

One of the more important procedural or administration questions involves the general problem of respondent anonymity. For most industrial uses, respondents to attitude scales or questionnaires need not be identified. It is the purpose of management usually to measure the *group* feeling; knowledge of the attitudes of specific individuals is not necessary in determining the over-all strengths and weaknesses leading to the need for administrative action. It is common practice, therefore, to emphasize during the scale administration that employees need not—indeed, that they should not—sign their names to their individual questionnaires. Often the administration procedure provides for a sealed ballot box into which employees are asked to drop their *unsigned* questionnaires. It would appear that these procedures would be sufficient to protect the anonymity of individual respondents. As a matter of fact, from the administrator's and from the firm's standpoint, it is sufficient.

However, a more important consideration has to do with the impression conveyed to employees who are completing the questionnaire. Hyman (2) has called attention to the necessity for distinguishing between *literal* anonymity and *psychological* anonymity. For example, questionnaire studies normally ask for such items of information as department number, age, length of work experience, etc. Requirements for this kind of information could easily suggest to an employee that special efforts are being taken to secure identifying information.

Furthermore, the employee's perception of the testing situation could have an important effect on his view of the use to which the attitude results might be put. Situational factors which could affect the general climate of psychological anonymity might include, among other things, the extent to which rapport is established between employees and

[1] The authors wish to acknowledge the assistance of Earl F. Cheit and Thelma Kunde who aided in the planning of the study and the administration of the questionnaires.

Table 1

Comparison of Sex, Age, Experience, and Education Characteristics of Samples A and B

Samples	Sex (% Female)	Experience (years)			Age (years)			Education (years)		
		Mean	*SD*	*t*	Mean	*SD*	*t*	Mean	*SD*	*t*
A	82%	4.3	4.8	1.09	40.1	17.4	0.56	10.6	2.1	1.00
B	82%	5.9	8.4		42.0	13.4		10.3	1.7	

administrator. In this vein, the actual identity of the test administrator could have an effect on responses obtained. It is this factor—the identity of the test administrator—that has prompted the present study.

The Industrial Relations Center (IRC) policy in the administration of its Employee Attitude Questionnaire (1, 3) has been to assign responsibility for scale administration to an IRC staff member or to some representative (not a member of the particular firm) appointed by the IRC executive staff. The reasoning behind this policy is based on the belief that every effort should be taken to accompany literal anonymity with psychological anonymity.

However, this policy is not universal. Many widely used standardized attitude scales do not provide, necessarily, for an "external" survey administrator. To be sure, questionnaires usually are unsigned and a ballot box is used, but it is rare that any effort is made to control possible effects resulting from "internal" administration of the scale.

Recently, the Industrial Relations Center was afforded opportunity to conduct a study directed toward discovering the extent of attitude-item response distortion attributable to the identity of the administrator.

It was decided to compare attitude scale responses obtained by two different survey administrators, the first an IRC staff member; the second, an official of the firm in which the survey was being conducted. The major hypothesis was that the presence of an official from the firm acting as the survey administrator would constitute more of a threat to employees' feelings of psychological anonymity than would the presence of an "external" administrator such as an IRC staff member.

From this major hypothesis, four subhypotheses were formulated for testing.

1. A threat to employees' feelings of psychological anonymity would result in their responding more favorably to the survey than employees not so threatened.
2. A threat to psychological anonymity would result in differential amounts of response distortion depending upon the content of different items comprising the questionnaire.
3. Employees with initially unfavorable attitudes would contribute most to any response distortion occurring under the threatening condition.
4. Employees feeling a threat to anonymity would tend to give fewer and shorter responses to open-end questions than employees not so threatened.

Method

The measuring instrument used in the study was the IRC Employee Attitude Questionnaire (1, 3). The scale employs Likert-type response categories. Scoring is done by either one of two methods: (*a*) the traditional Likert technique of assigning 0, 1, 2, 3, or 4 points to each response depending on its degree of favorableness, or (*b*) by reporting the percentages of responses given in the favorable, unfavorable, or undecided categories. Attitudes measured by the 53 items of the questionnaire are divided among seven subscales which have been labeled: General Morale, Working Conditions, Type of Work, Hours and Pay, Supervision, Co-Workers, and Communications. The number of items in these scales varies from five in the Communications and Supervision areas to 16 in the General Morale Scale.

In addition to these seven standardized scales, three open-end questions are included which read as follows:

1. What do you like best about working for this Company?
2. If you were made head of this Company, what would you do to make this a more pleasant place to work?
3. If you have any more comments to make about your job, your supervisor, or the Company, write them below.

The study was undertaken in a large department store in the Twin Cities area. A list of all rank-

and-file employees was obtained from the personnel department. Samples were drawn randomly from this list to provide subsamples of 45 individuals each.

An Industrial Relations Center staff member administered the Attitude Questionnaire to persons in Sample A. The personnel manager (known to all employees) of the department store administered the questionnaires to persons in Sample B. The methods of administration were identical in all other respects. Employees were assured that their replies were confidential; they were specifically asked not to sign their names, and they placed their completed questionnaires in a ballot box which was labeled PROPERTY OF THE UNIVERSITY OF MINNESOTA.

As a check on the comparability of the samples, information concerning sex, age, length of experience, and education was obtained. Comparisons based on these four factors yielded the data shown in Table 1.

Information in Table 1 shows that none of the differences between samples is sufficiently large to reject the hypothesis that the samples are drawn from the same population.

Results

Table 2 shows mean scores, standard deviations, and *t* values for each of the samples on each of the subscales of the Employee Attitude Questionaire.

Mean scores for Sample B (the internally administered group) are higher (i.e., more favorable) on every subscale than the mean scores of Sample A. The differences are statistically significant on two of the seven subscales: General Morale and Supervision. In terms of overlap, from 30% to 35% of employees in Sample A exceed the median of those in Sample B. It appears that, for rank-and-file employees at least, a moderate degree of distortion has occurred in the expected direction. The first hypothesis is, therefore, supported.

In order to test the second hypothesis, the one concerning differential effects of item content, seven IRC staff members were asked to identify items the content of which might appear to be particularly threatening to an employee who was not completely convinced of the anonymity of his responses. Thirteen items, on which five or more of the staff members agreed, were selected as comprising such a "Threat Scale." These items include statements directed primarily toward interpersonal and supervisory relationships with heavy emphasis on the degree of understanding extended toward employees and the fairness with which employees are treated.

Table 2

Summary Statistics for Attitude Scores Obtained by Samples A and B of Rank-and-File Employees

		Sample A IRC-Administered ($N=45$)		Sample B Internally Administered ($N=45$)
General morale	Mean	37.9		42.5
	SD	9.5		9.4
	t		2.28*	
Type of work	Mean	17.1		17.6
	SD	3.0		3.4
	t		0.76	
Co-workers	Mean	13.5		14.0
	SD	3.5		3.2
	t		0.81	
Working conditions	Mean	18.9		19.7
	SD	5.0		4.8
	t		0.73	
Hours and pay	Mean	18.6		19.4
	SD	3.5		4.2
	t		0.90	
Supervision	Mean	13.6		15.2
	SD	3.9		3.4
	t		2.01*	
Communications	Mean	11.9		12.3
	SD	2.9		3.3
	t		0.71	
Total score	Mean	131.3		140.7
	SD	22.8		23.9
	t		1.90	

* Significant at the 5% level.

The IRC staff members also selected items which they believed definitely did not include "threatening" content. There was good agreement on 32 such items.

Table 3 shows comparisons between mean scores obtained by the two groups on the Threat and Nonthreat Scales. It is seen that the major point at which distortion of responses occurs is on the Threat Scale items. In terms of overlap, only 33% of Sample A rank-and-file employees exceed the median of employees in Sample B. The range of scores for the former group is 3 to 50 and that for the latter group, 22 to 52. Analysis of responses to single items shows that the group to whom the personnel manager administered the survey gave a larger percentage of

Table 3

Summary Statistics for Scores Obtained by Employee Groups on Threat and Nonthreat Scales

	Threat Scale		Nonthreat Scale	
	Sample A	Sample B	Sample A	Sample B
	IRC	Internal	IRC	Internal
Mean	32.3	37.3	79.5	82.5
SD	9.4	7.5	11.4	13.9
t	2.78**		1.13	

** Significant at the 1% level.

favorable responses to every item in the threat scale. The percentage differences range from a low of 2% to a high of 25%, with a median of 13%. Differences in percentage of favorable responses on the Nonthreat items were about equally divided between positive and negative values with a median of only 1%.

These data support the second hypothesis. The effect of the identity of the administrator upon survey responses is most marked among certain definite items whose content involves attitudes toward various aspects of supervisory behavior and the spirit of fair play in the work setting.

The third hypothesis has to do with the extent to which different individuals with different initial attitudes might be prone to give responses that are more favorable than warranted by their actual or "true" attitudes. Two methods were used in an effort to test this hypothesis.

First, employees in the two groups were ranked according to their score on the Threat Scale. Employees with corresponding ranks were paired. The ranks were normalized and the resulting scale scores were correlated with differences between threat scores for the 45 employee pairs. The correlation between these variables is $-.57$. If the two extreme values at the lower end of the continuum are deleted, however, the correlation is reduced to $-.41$. These negative correlations suggest that differences between the two groups tend to be less at the more favorable or higher end of the continuum. However, examination of the scatter diagram suggested that correlations were raised spuriously by a small clustering of 8 or 10 measures in the middle portion of the continuum. A second method of analysis was therefore undertaken.

Rank-and-file employees scoring in the top third and bottom third of the total Attitude Questionnaire were designated, respectively, as having generally favorable and generally unfavorable attitudes toward their total work situation. Comparisons were then made between Threat scores obtained by high-scoring persons in Samples A and B and between low-scoring persons in Samples A and B. The results are shown in Table 4. Both comparisons result in differences significant at the 5% level. The mean difference between the favorable groups is 3.6 points as compared to a mean difference of 5.6 points for the unfavorable groups. At first glance, the "larger" difference between the unfavorable groups appears to lend support to the hypothesis being tested. However, these figures should be interpreted in terms of the variabilities of scores at the two points on the attitude continuum. In other words, we need to examine the extent of overlap between the sets of distributions. In terms of overlap, about 13% of the employees in *each* of the two IRC-administered samples exceed the medians of the firm-administered samples. There is no difference, therefore, between the amounts of response distortion shown by initially favorable and initially unfavorable employees. It appears that evidence is not sufficient to support a claim that employees with unfavorable "true" attitudes will distort or shift their responses to a greater extent than employees with favorable "true" attitudes.

Table 4

Summary Statistics for Threat Scores Obtained by Employees in Samples A and B Who Have Favorable and Unfavorable Attitudes

	Favorable Over-all Attitudes		Unfavorable Over-all Attitudes	
	Sample A IRC ($N=16$)	Sample B Internal ($N=15$)	Sample A IRC ($N=15$)	Sample B Internal ($N=15$)
Mean	40.9	44.5	23.7	29.3
SD	3.7	3.7	8.4	4.7
	2.59*		2.18*	

* Significant at the 5% level.

The final hypothesis involved the incidence of "open-end" responses among the two samples. This effect was determined simply by counting the number of employees choosing to respond to each of the three open-end questions. Results are shown in Table 5.

In every case, a larger proportion of persons in Sample A (the IRC-administered group) responded to the open-end questions than in Sample B. All of the differences are statistically significant. Furthermore, the average length of each comment given by the former group was 37 words as compared with an average of only 24 words for each comment by the latter group. The effects of these two factors combined with the result that the IRC-administered group produced exactly twice as much verbal material as the group to whom the firm's personnel manager administered the questionnaire. This is striking support in favor of the fourth and final hypothesis. It appears that the identity of the administrator can have an important effect on the desire of employees to respond freely to open-end questions.

It has been the IRC's experience that management officials in the firms regard the verbatim comments from their employees as especially valuable supplementary evidence to the scores obtained on the standardized scales. In view of this, the finding with regard to willingness to answer open-end questions may well be the most relevant of the entire study.

This study has led the IRC staff to conclude, for the time being at least, that the policy of using staff members or staff representatives to administer the Triple Audit Questionnaire is well founded and that it should be continued. It appears that the necessity for distinguishing between literal anonymity and psychological anonymity has been confirmed, and a few steps have been taken in defining the attitude areas which may be most susceptible to the subtle effects of different survey administrators. The study suggests, particularly, that any survey directed toward securing employee attitude information bearing on interpersonal and supervisory relationships should take special pains to engender feelings of complete psychological anonymity on the part of the employees participating in the survey.

Table 5

Proportions of Persons Responding to Open-End Questions

Samples	What do you like best about working here? (% responding)	What would you do if you were made head of this company? (% responding)	Other Comments (% responding)
A	76%	69%	44%
B	49%	49%	16%
	$t = 2.76^{**}$	$t = 1.98^{*}$	$t = 3.04^{**}$

* Significant at the 5% level.
** Significant at the 1% level.

Summary and Conclusions

A study was conducted to discover the effects on attitude scale responses of the identity of the survey administrator. Two employee samples were selected randomly from the total work force of a large Twin Cities department store. An Industrial Relations Center staff member administered the IRC Employee Attitude Questionnaire to one of the groups; the personnel manager of the store administered the questionnaire to the other group.

The results gave support to the following hypotheses:

1. A threat to employees' feelings of anonymity results in their responding more favorably to an attitude survey than employees not so threatened.

2. A threat to anonymity results in differential amounts of response distortion depending upon the content of different items comprising the questionnaire.

3. Employees feeling a threat to anonymity tend to give fewer and shorter responses to open-end questions than employees not so threatened.

Received June 10, 1955.

References

1. Heneman, H. G., Jr., & Yoder, D. Employee opinion survey by remote control. *Personnel J.*, 1953, **32**, 169–172.
2. Hyman, H. H. *Interviewing and social research.* Chicago: Univer. of Chicago Press, 1954.
3. Yoder, D., Heneman, H. G., Jr., & Fox, H. *Auditing your manpower management.* Minneapolis: Univer. of Minnesota Press, May, 1954.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Attitudes of Union and Non-Union Employees in a Calcutta Electrical Engineering Factory

Harish Chandra Ganguli

Indian Institute of Technology, Khargpur

Every worker in a factory is subject to cross pressures for loyalty to the two membership groups—the company and the labor union. It is also to be expected that if a worker is a member of the union, he perceives, judges and feels about things in the work situation in which the union is interested in a different way than the worker who has not joined the union. Largely because of this influence of the group on the norms and attitudes of its members, the morale of the worker, meaning thereby his contentment with the employment relationship, is also likely to be affected by this union membership.

The Study

A morale survey was made by the author in an important engineering factory in Calcutta which produced all types of electric fans. In 1951 it had 1,890 workers on its roll and produced 62,188 electric fans which is equivalent to 29.3 per cent of the country's total fan production. The factory had a labor union controlled by extreme leftist trade unionists since 1945. This union (to be referred to as the outside union or simply the union) is not recognized by the company although it has a substantial hold over the workers, having fought several successful and partly successful battles for them in the past. In 1952 about 55 per cent of the workers were members. The company has tried to help the development of a rival union in the factory which is commonly known amongst the workers as the "inside" union. Its membership is limited, however, to about 8 per cent of the workers, the rest being unattached to either of these two organizations.

The morale survey involved all workers from the foundry shop, the machine shop, and one assembly shop, totaling 550 workers. The method of study closely resembles the sample interview survey technique (**6**), though on a much smaller scale. An attitude scale on the principle of summated ratings (**5**) was constructed on the basis of initial exploratory interviews with 40 workers and four pretests. The data regarding satisfaction of the workers were obtained from systematic interviews with them on the basis of this scale. The attitudinal items had fixed alternative answers and the relevant response alternatives were checked by the interviewer during a free-answer interview with each worker. Each interview took between one and one-and-a-half hours to complete. In the construction of the scale and conduct of the interviews the usual precautions were taken for ensuring the validity of the results (**1**, **7**, **9**).

Results

The preliminary scale contained 41 items. To know the primary morale dimensions that were covered by these items, a factor analysis by Thurstone's centroid method was made with 23 of the best items in the scale. These items were selected on the basis of discrimination values and item analysis. The factor analysis led to the isolation of three factors or dimensions of morale. The first factor, Factor C, refers to the worker's satisfaction with the benefits he derives from the employment relationship and also his over-all confidence and satisfaction with the total organization. It refers to issues like how the factory compares with other factories as a place to work, if the company is generally sympathetic to the workers and appreciative of their point of view, and also to specific issues like income, the welfare activities of the company, the effectiveness of supervision, etc. Factor So gives a measure of the workers' satisfaction with the technical and organizational aspects of supervision, and has reference to issues like division of work load, attention given to their suggestions regarding methods of work, tools, etc., providing ade-

Table 1

Distribution of Scores of 548 Workers on the Three Morale Factors

Score	Factor C (No. of Workers)	Score	Factor So (No. of Workers)	Factor Sp (No. of Workers)
8–12	104	5–7	24	39
13–17	266	8–10	103	113
18–22	132	11–13	158	193
23–27	35	14–16	154	181
28–32	11	17–19	109	22

Table 2

Percentage of Workers Obtaining Different Scores on Factor C with Different Union Groups Treated Separately

Score	No. of Workers in Percentage		
	Inside	Non-Union	Outside
8–12	6.5	10.8	26.5
13–17	26.1	44.6	54.7
18–22	41.3	30.4	17.1
23–27	15.2	11.3	1.7
28–32	10.9	2.9	0.0

quate facilities for the men to do their work properly, and so on. Factor Sp refers to the satisfaction the worker derives from his relations with the supervisor as a person. The supervisor's skill in handling his men, the reasonableness of what he expects from them, and, in general, whether he can be regarded as their "own man" are important points in this connection.

For each worker, therefore, measures for these three types of satisfaction have been obtained. Table 1 gives the distribution of scores on these factors for the 548 workers.[1]

For the three union groups separately, Tables 2 and 3 give the percentage of workers in each score interval in Factors C, So, and Sp.

On analysis it was found that in each morale dimension the outside union group was most dissatisfied, the inside union group was most satisfied, and the non-union group occupied a middle position between these. This is seen from Table 4. Also t tests show that the difference in satisfaction between any two union groups on each morale factor was significant to the 1 per cent level except for the difference between inside union and non-union groups in Factor So which is significant at the 5 per cent level only.

To find out if this relation between membership of the unions and the satisfaction of the workers was genuine, further analysis was made of this relationship by controlling on other variables that have influenced worker satisfaction. The variables controlled were supervisory group, nature of work, income, pay increase in last three years, financial aspiration, length of service, schooling of the workers, etc. Thus, for example, relation between union membership and confidence in the company (Factor C) has been determined by statistically comparing outside and non-union workers (inside union workers were too few) under the same supervisor, doing the same type of work, and in the same income group. Such analyses showed that in almost every case the inside union workers were most satisfied and outside workers least satisfied, and the non-union workers neither very satisfied nor very dissatisfied. With regard to attitude toward supervision, it was also noted that in comparatively ill-managed shops the discrepancy between attitudes of union and non-union workers was greater. The effectiveness of supervision thus seemed to act as an intervening variable.

A further comparison of the different union groups has been made regarding the extent of their dissatisfaction in general. Owing to shortage of space, however, differences over specific issues are not discussed. Such an over-all comparison of these union groups from the same factory is to some extent justified since they were subject to the same type of company policies and practices and the same personnel policies. These groups were also comparable to some extent regarding the work they did and the general quality of su-

[1] Data for two workers were incomplete and therefore ignored. Because of lack of background data for six others, most statistics from Table 4 onwards have been calculated on the following frequencies: Inside Union—46, Non-Union—202; Outside Union—294. These samples are roughly proportionate to the membership of the three union groups in the whole factory.

Table 3

Percentage of Workers Obtaining Different Scores on Factors So and Sp with the Different Union Groups Treated Separately

Score	Factor So			Factor Sp		
	No. of Workers—%			No. of Workers—%		
	Inside	Non-Union	Outside	Inside	Non-Union	Outside
5–7	—	1.0	7.4	—	2.0	11.7
8–10	8.7	6.4	28.9	6.5	13.7	27.5
11–13	8.7	26.5	33.5	21.7	33.8	38.3
14–16	34.8	37.2	20.8	58.7	46.1	20.1
17–19	47.8	28.9	9.4	13.0	4.4	2.4

pervision, since in each of the three shops—foundry, machineshop, and the assembly lines—there were some inside union workers, some outside union workers, and some non-union workers.

Assuming that attitude was normally distributed, the range of morale scores along each factor was divided into five segments of 1.2 σ each. These five subgroups may be regarded as categories of satisfaction, ranging from the "most dissatisfied" category to the "most satisfied" category. Table 5 below gives the percentage of workers in each satisfaction group, union-wise.

Table 5 shows that always there were more outside union workers than non-union workers in the dissatisfied groups. Inside union workers, on the other hand, were more satisfied than the other two groups. Also the percentage of satisfied and dissatisfied workers in any union group over all the three morale dimensions were similar. About 40 per cent of outside union workers were dissatisfied over each morale dimension, and about 40 per cent of non-union workers and 60 per cent of inside union workers were satisfied.

Table 4

Mean and Standard Deviations of Scores on the Three Factors for the Three Union Groups Treated Separately

Group	Factor So	Factor Sp	Factor C
Inside union	15.5±2.62	14.4±2.13	19.9±5.27
Non-union	14.4±2.58	13.2±2.48	17.6±4.68
Outside union	11.8±2.99	11.3±2.35	14.7±3.54

Discussion

This study was originally intended to determine variations in industrial morale and factors responsible for these variations. Very soon, however, it was realized that the union membership of some workers and nonattachment to it of others was the most important single factor related to their job contentment. Significantly negative biserial correlations between the three morale factors and "outside union membership—no union membership" were found. These are

With confidence in company, etc. (Factor C)	−0.410 ±0.049
With technical aspects of supervision (Factor So)	−0.500 ±0.045
With human relations aspects of supervision (Factor Sp)	−0.396 ±0.049

This inverse relation was found to exist even when analysis was made after keeping constant other variables like nature of supervision, nature of work, etc. An over-all view of this influence is shown by the fact that about 60 per cent of inside union workers and 40 per cent of non-union workers in contrast to only 12 per cent of outside union employees expressed satisfaction in the three areas of their employment relationship. An analysis of attitudes on individual issues showed that maximal difference in satisfaction between the outside union group and the other two groups existed in connection with (*a*) relation with one's supervisors and allocation of the work load, and (*b*) the treat-

Table 5

Percentage of Outside, Inside, and Non-Union Employees in Different Satisfaction Subgroups

Area	Group	Most Dissatisfied	Dissatisfied	Neutral	Satisfied	Most Satisfied
Technical aspects of supervision (Factor So)	Outside union	6.0	35.6	45.7	12.1	0.6
	Non-union	0.5	10.3	48.0	35.3	5.9
	Inside union	0.0	8.7	30.4	50.0	10.9
Human relations aspect of supervision (Factor Sp)	Outside union	5.0	33.9	48.3	10.8	2.0
	Non-union	2.0	13.2	43.1	37.7	4.0
	Inside union	0.0	6.5	34.8	47.8	10.9
Satisfaction from wages, etc. and with the total organization (Factor C)	Outside union	4.7	32.6	49.0	13.4	0.3
	Non-union	2.0	15.2	43.1	34.8	4.9
	Inside union	2.2	6.5	30.4	43.5	17.4

ment received from the company and how this factory compared with other similar factories as a place to work.

It is not known and cannot be exactly known from the present study how far the union (outside) is responsible for this greater dissatisfaction among its members. That it made a significant contribution to such a condition may be inferred from the enormous efforts it made by means of mass rallies, pamphlets, wall placards, through informal group discussions and on a national level, through its newspaper to spread dissatisfaction among its workers with their wages and conditions of service and with the generally capitalistic and antiworker attitude and policies and practices of the management. But it was also true that many workers who join the union, especially in its early formative stage, are somewhat more dissatisfied and anticompany in their attitude than the average. Therefore this relation between poor morale of the worker and his union membership seemed to be more correctly expressed as a circular causal relation. Each was to some extent the cause of the other, and in turn is also affected by the other.

That the results of the present study are not altogether peculiar to the specific plant is supported by a similar study conducted by the author in another important light engineering factory in Calcutta (**2**, **3**). This factory had only one union, which was also Communist-dominated and not recognized by the company. The results showed that whereas 38 per cent of non-union employees were generally satisfied with their employment relationship, only 23 per cent of the union members were satisfied. Analysis showed this negative relation to be present even when factors like supervision, earnings, and education of workers were kept constant. Attitudes between the two groups over specific issues varied in the same way as for workers in the present study.

Union membership and poor morale need not be associated necessarily. In a study made in the United States and reported by Kolstad (**4**), for example, it was found that there were very few significant differences when the attitudes of union and non-union employees were compared. One difference was regarding pay, an item covered by union contract. Fifty-two per cent of non-union workers as compared with only 30 per cent of union workers were dissatisfied with it, and also 71 per cent as compared to 42 per cent of non-union employees thought that they got as high or higher pay than they would get for similar work in other factories. Non-union employees as a group, however, felt more definitely that they "belonged," that they were part of the organization.

The important factors in this context seem to be the attitude of management toward labor organizations, the policies and practices of these organizations and the resulting de-

gree of cooperation between the two. In India many managements do not feel that labor has any right or grounds to organize as evidenced by the many unrecognized unions. Many unions, especially those affiliated with some extreme leftist organizations, do not always follow a policy dictated by the workers' interests alone and thus cooperation between labor and management in the form of collective bargaining, etc. is conspicuous only by its absence. The degree to which these characteristics are present in any factory determines the association between union membership and poor worker morale, and they were largely present in the two factories studied; hence the results as noted above. Incidentally, the very large difference of opinion between union and non-union employees over aspects of supervision emphasizes its importance for the worker as pointed out by Roethlisberger (8), and also points to the possibility that many supervisors adopt management goals as opposed to union goals for their frame of reference. Again, the substantial difference in the way these two groups evaluate general issues like the company's treatment of and sympathy for the worker, its wage payment policy, and absence of any such difference over specific issues like satisfaction with nature of work done, wages received, and so on seem to be another illustration of the principle that social influences are more powerful in ambiguous and vague situations than in a clearly structured situation.

Received June 14, 1955.

References

1. Anonymous. *Human relations study techniques.* Ann Arbor: Survey Research Center, Institute for Social Research, 1949.
2. Ganguli, H. C. A study on effect of union membership on industrial morale. *Indian J. Psychol.,* 1954, **29**, 45–60.
3. Ganguli, H. C. Relation of union membership to attitude of industrial workers. *Indian J. soc. Wk,* 1954, **15**, 189–199.
4. Kolstad, A. Excerpts from employee attitude survey. In M. L. Blum (Ed.), *Readings in experimental industrial psychology.* New York: Prentice-Hall, 1952, pp. 114–117.
5. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.,* 1932, No. 140.
6. Likert, R. The sample interview survey. In W. Dennis (Ed.), *Current trends in psychology.* Pittsburgh: Univer. of Pittsburgh Press, 1947. Pp. 196–225.
7. Marriott, R. Some problems in attitude survey methodology. *Occup. Psychol., Lond.,* 1953, **27**, 117–127.
8. Roethlisberger, F. J. The foreman: master and victim of double talk. In S. D. Hoslett (Ed.), *Human factors in management.* London: Macdonald & Evans, 1948. Pp. 51–72.
9. Vernon, P. E. The assessment of psychological qualities by verbal methods. *Industr. Hlth Res. Bd. (Lond.),* 1952, Rep. No. 83.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Prediction of Success from Examination of Performance During the Training Period

Patricia C. Smith and Robert A. Gold

Cornell University

At what point in the learning period can ultimate success on the job be predicted? This question arises whenever probationary periods must be set, or whenever a foreman must decide whether a particular learner should continue training, be transferred, or be dropped. Frequently the problem is whether actual performance for a given week is a better predictor of eventual success than a battery of tests used in selection.

There is evidence that the abilities important in job performance after the end of the learning period are frequently different from those of the learning period. Investigations of the predictive value of tests for trainability and job proficiency show that shifts in aptitude requirements may take place. For example, Brown and Ghiselli (2) correlated pairs of validity coefficients obtained from a number of studies where both job proficiency and trainability criteria had been used in validating the same kinds of tests for similar jobs. Correlations between validity coefficients using the two criteria were low. This was surprisingly true even for clerical jobs where verbal, numerical, perceptual, and reasoning abilities could be expected to be important during and after learning. Evidence of a more direct nature of the relationship of early performance to eventual achievement is found in Kornhauser's study of billing-machine operators for whom the shift could be expected to be small (4). Even so, his correlations were not high, especially for early parts of the learning period.

The most desirable criterion would be production after completion of learning, but previous studies have used either time to reach average production (1, 5) or total of production during training (1). Even with these criteria that reflect performance during learning, prediction based on the early weeks of the learning period did not prove to be very accurate. Table 1 summarizes the results of these earlier studies.

The following hypotheses concerning the correlations between production at various stages of learning and eventual productivity for power sewing-machine operators were tested in the present study:

1. They are low during the first weeks of initial learning, when the learning curve is rising most rapidly.
2. They rise steadily throughout the learning period.
3. They are lower than those found in other studies in which performance during learning entered into the criterion figures.
4. They are lower than those obtained for clerical workers, for whom the abilities required during learning may be expected to be more similar to those required after training.

Procedure

Production data were gathered for the first 22 weeks of employment of 18 women, previously untrained in power sewing-machine operation. They hemmed handkerchiefs, using specialized folder attachments. All operations were nearly identical. Payment was by piece rate, with a guaranteed minimum. Employment was seasonal, and the workers were employed over almost exactly the same period of time. There had been almost no selection, and there were very few terminations prior to the end of the season, so that the range is unusually unrestricted. All learning curves had leveled by the end of the eighteenth week. There was no increase in the average learning curve after Week 19, and in only one case was production for Weeks 21 or 22 more than one per cent greater than that for Weeks 19 or 20. In this exception, a difference of 8 per cent, production after Weeks 21 and 22 dropped back to the level of Weeks 19 and 20. The average curve showed that the first sharp rise was completed by Week 5.

Average hourly production for Weeks 19 through 22 was taken as a criterion of performance after learning. Rank correlations were computed between this criterion and figures for each week from Week 3, when complete production records were first avail-

Table 1

Correlations Between Production at Various Stages of Learning and Different Criteria of Proficiency after Learning, Previous Studies

Job Studied	Investigator	Learning Figure Used	Weeks to End of Learning Period	Percentage of Learning Period Elapsed	Correlation with Criterion of Proficiency after Learning: Time to Reach Average Production	Total Production During Training	Speed at End of Learning Period	Accuracy at End of Learning Period
Preparation of spools for rug looms	McGehee (5)	Average production for single week	46	2%	.599			
				4%	.836			
				7%	.760			
				9%	.792			
				11%	.833			
				13%	.629			
				15%	.842			
				17%	.679			
Billing-machine operators	Kornhauser (4)	Speed of operation	26*	17%			.20	
				33%			.48	
				50%			.60	
				67%			.76	
				83%			.76	
Billing-machine operators	Kornhauser (4)	Accuracy of operation	26*	17%				.00
				33%				.25
				50%				.31
				67%				.58
				83%				.74
Preparation of spools for rug looms	McGehee (5)	Average of cumulative production through week	46	2%	.599			
				4%	.834			
				7%	.842			
				9%	.887			
				11%	.894			
				13%	.929			
				15%	.924			
				17%	.923			
Power sewing-machine operation:	Blankenship and Taylor (1)	Average of cumulative production through week						
covering			50**	20%	.41	.01		
hemming			40**	25%	.32	.66		
trimming			30**	33%	.66	.68		

* Estimated.
** Estimated from learning curves for both workers completing learning period and those not completing, and from reported difficulty of job.

able, through 10, when half the learning period had been completed and the average learning curve had reached 85 per cent of the terminal production rate. Average production for each two-week period and for each three-week period during the first 10 weeks was also correlated with the criterion.

Results and Discussion

Table 2 shows a summary of results. The first statistically significant correlations are for Week 5, or for composites including Week 5. This was the point at which the

initial rise of the learning curve had been completed. The foreman who wishes to make his decision on the basis of a single week's production, or the average of two or three weeks' production, is probably justified in doing so after the completion of a fourth of the learning period, since correlations in this study, as in all previous studies (see Table 1), reach a significant level by that time. Furthermore, the correlations at that point probably exceed those of most test batteries with eventual production. Decisions based on cumulative data may, in some cases, be warranted slightly prior to that point, but caution should be exercised when less than 20 per cent of the learning period has elapsed, particularly if there is any contradictory evidence to suggest a delay, or if training has been expensive.

The correlations in this, as in preceding studies, rise steadily with time, suggesting that there are more factors in common in the later portions of the learning period.

The correlations are substantially higher than those for Kornhauser's clerical workers, and lower than those of McGehee and of Blankenship and Taylor for factory workers. These differences are probably too great to be attributable to errors in estimation of lengths of learning periods. There is no indication of greater change in the ability patterns involved in learning motor skills than in acquiring clerical skills. The obtained hierarchy might better be accounted for by differences in the extent to which learning speed appears in the various criteria. Since Kornhauser's criteria were further removed from the training period than ours, and ours are correspondingly less contaminated with effects of learning speed than those of the other two studies, it seems likely that correlations as low as Kornhauser's might be obtained in the other situations after a longer lapse of time.

The lack of evidence for a lower relationship between learning and job performance scores for motor than for clerical workers is in line with the studies of validity by Brown and Ghiselli (2), who found correlations between validities for trainability and job proficiencies to be equally low for clerical as for other kinds of jobs. Patterns of ability required appear to shift to as great extent in one situation as in the other.

Table 2

Correlations Between Production at Various Stages of Learning and Average Production at End of Learning Period, for Hemming Handkerchiefs

Learning Figure Used	Weeks	Percentage of 20.5 Weeks Learning Time Elapsed	Rank Correlation with Average at End of Learning Period
Single week	3	15%	46
	4	20%	40
	5	24%	55*
	6	29%	73**
	7	34%	74**
	8	39%	78**
	9	44%	81**
	10	49%	80**
Average, two weeks	3, 4	20%	46
	4, 5	24%	48*
	5, 6	29%	67**
	6, 7	34%	73**
	7, 8	39%	79**
	8, 9	44%	81**
	9, 10	49%	82**
Average, three weeks	3, 4, 5	24%	48*
	4, 5, 6	29%	60**
	5, 6, 7	34%	68**
	6, 7, 8	39%	78**
	7, 8, 9	44%	81**
	8, 9, 10	49%	83**

* Rho significant at 5% level (3).
** Rho significant at 1% level (3).

Summary and Conclusions

Production during early weeks of the learning period was correlated with average production after the completion of learning for a group of power sewing-machine operators. Four hypotheses were proposed concerning these correlations; three proved tenable:

1. Correlations are low during the first few weeks of training, when the learning curve is rising most rapidly.

2. Correlations rise steadily as the learning period progresses.

3. Correlations are lower than those obtained in previous studies using criteria which reflect performance during learning rather than ultimate level of proficiency.

A fourth hypothesis received no support in the present study. Correlations were not lower but, in fact, somewhat higher than those obtained previously for clerical workers, for whom it was suggested that abilities required during learning might be more similar to those required later than for the perceptual-motor tasks.

Received June 24, 1955.

References

1. Blankenship, A. B., & Taylor, H. R. Prediction of vocational proficiency in three machine operations. *J. appl. Psychol.*, 1938, **22**, 518–526.
2. Brown, C. W., & Ghiselli, E. E. The relationship between the predictive power of aptitude tests for trainability and for job proficiency. *J. appl. Psychol.*, 1952, **36**, 370–372.
3. Kendall, M. G. *The advanced theory of statistics.* London: Charles Griffin, 1948. Vol. 1, p. 401.
4. Kornhauser, A. W. A statistical study of a group of specialized office workers. *J. pers. Res.*, 1923, **2**, 103–123.
5. McGehee, W. Cutting training waste. *Personnel Psychol.*, 1948, **1**, 331–340.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

The Measurement of Personal Factors Related to Success of Office Workers [1]

Mearl R. Guthrie

Bowling Green State University

Numerous studies have shown that the personal traits of employees are considered to be a very important factor by well-informed employers. Research reveals that undesirable personal characteristics are the most common reason for the failure of office workers to be promoted. Various attempts have been made to measure the personal traits needed for success on the job, but there still remains a definite need for an easy-to-interpret technique that will aid in the selection of employees for office work on the basis of their personal traits.

Problem

The main purpose of this study was to develop a technique which would aid in the differentiation of employees who have desirable personal traits for office work from those who do not possess such personal traits. The secondary purposes were to determine the validity and reliability of the technique after it had been developed.

The study was limited to female office workers whose duties did not involve supervising or order giving. It was further limited to offices in which five or more employees worked together, because most office executives agree that the ability to work and get along with other people is one of the most important traits an office worker can have.

A definition of all the personal traits necessary for success in office work was not attempted. It was considered more appropriate to develop and validate the instrument on the basis of over-all personal characteristics for office work.

Procedure

Development of the items. Several types of instruments and items were considered, and the opinion type survey form was chosen for the purpose at hand. A sample of the opinion-type statement is shown below:

The supervisor is usually to blame when employees fail to follow directions. SA a ? d SD

After reading such a statement and deciding how one feels about it, the office worker draws a circle around one of five possible answers—Strongly Agree, agree, uncertain or undecided, disagree, or Strongly Disagree. A total of 457 items were developed by a canvass of the literature and an evaluation of the experience of a number of people with a variety of backgrounds, in office work, psychology, testing, etc.

Selection of items. It is a relatively easy task for an office executive or supervisor to select office workers at the extreme ends of the personal trait continuum. The officials of 50 offices, representing a wide variety of business activities, were asked to select one or more employees who had very satisfactory personal traits for office work and one or more employees who had very unsatisfactory personal traits for office work. Specific criteria were provided for each official to aid him in selecting the two extreme groups.

The Survey, consisting of the 457 items, was administered by the writer under a standard set of conditions and instructions to 100 "satisfactory" and 100 "unsatisfactory" office workers. It was hoped to minimize the possibility of the employee faking his opinions by calling the instrument a survey and by not requiring the employee to identify his paper. However, various methods were used to identify the papers as satisfactory or unsatisfactory after they were completed.

Each item on each survey was tallied on a tabulation sheet and the totals in each of the cells computed. These totals were also percentages because there were 100 papers in each of the two groups. The following item is an example:

"Those in charge" have much more freedom in the office than the employees.

	SA	a	?	d	SD
Satisfactory	12	39	9	35	5
Unsatisfactory	8	62	3	25	2

The following criteria were used in the item-selection process:

1. Did the item adequately differentiate between the two groups of office workers?

2. Were the differences in percentages consistently positive or negative in the Strongly Agree and agree categories, and in the disagree and Strongly Disagree categories?

[1] A dissertation submitted to the University of Minnesota in partial fulfillment of the Ph D. degree.

3. Was there at least one category in the item in which the difference between percentages was significant at the 10% level or less?
4. Was the item clear in meaning?
5. Did the statistic chi square substantiate the ability of the item to differentiate between the two groups?

A total of 74 items were chosen by this process as outstanding in their ability to discriminate between the two groups.

Scoring the items. A scoring key was developed by assigning weights of plus one, zero, and minus one to the five responses of each of the 74 items. These weights were determined by utilizing the difference between percentages in each of the five categories and the direction of the difference, as, for example, in the following item:

	SA	a	?	d	SD
Satisfactory	10	22	8	46	14
Unsatisfactory	16	48	8	26	2
Difference	−6	−26	0	20	12
Significance		1%		5%	1%
Weight		−1		+1	+1

Table 1 shows a comparison of the scores made by the two groups. A significant difference between the means of the two groups would be expected because these two groups were the basis for the selection of the items.

Selection of additional items. On the basis of the scores made on the 74 items, new criterion groups were established consisting of the upper and lower 25% of the original 200 surveys. The responses to the 457 items by the two new criterion groups were then tabulated. The same procedure used to select the 74 items was used to select a total of 150 items. Chi square was not used in the selection of additional items because it was found to add little, if any, objective information to the original item-selection process.

The comparison in Table 2 indicates that the scoring key developed for the 150 items was satisfactory. It is of interest to note that not one score of the lower 25% group exceeded the mean of the upper 25% group, and vice versa.

Table 1

Comparison of Scores Made by the "Satisfactory" and "Unsatisfactory" Groups of Office Workers on 74 Items

	Satisfactory	Unsatisfactory
N	100	100
High score	34	19
Low score	−13	−36
Range	47	55
Mean	12.75	− 7.15
SD	9.92	9.59

Validity

The method used in developing the items and the item-selection process contributed to the validity of the individual items. The procedure designed to determine the validity of the instrument as a whole was to (*a*) administer the 150 items to two unselected groups of office workers consisting of 73 full-time office workers and 54 business administration co-op students, (*b*) obtain ratings from several sources concerning the personal traits of each office worker, and (*c*) compute coefficients of correlation between scores and ratings.

Ratings for each individual worker were obtained from the following three sources: (*a*) the several fellow workers in each group, (*b*) the immediate on-the-job supervisor of each individual, and (*c*) the office manager or personnel manager of each company concerned for each full-time office worker and the University coordinator for each co-op student.

Ratings by fellow workers. An indirect method was developed by which each individual could be rated by the other members of her working group. A number of questions similar to the one below were used.

If I could have my lunch hour at the same time as one of the above persons,

My first choice would be _____
My second choice would be _____
My third choice would be _____
My fourth choice would be _____

Following completion of the Office Workers Opinion Survey, each individual was asked to answer seven questions on the Fellow Worker's Rating Scale concerning the other members of her group. These ratings were made

Table 2

Comparison of Scores Made by the Upper and Lower 25% of the Original Groups on 150 Items

	Satisfactory	Unsatisfactory
N	50	50
High score	83	− 14
Low score	− 31	−107
Range	114	93
Mean	23.1	− 58.3
SD	29	21.1

Table 3

Correlations Obtained Between OWOS Scores and Various Ratings on the Unselected Group of 73 Full-Time Office Workers

Ratings	2	3	4	Multiple Correlations
1. Office Workers Opinion Survey	.29*	.47**	.20	.51** (OWOS and 2, 3, 4)
2. Fellow workers' ratings		.43**	.55**	.49** (OWOS and 2, 3)
3. Supervisors' man-to-man ratings	.43**		.55**	.30* (OWOS and 2, 4)
4. Office managers' man-to-man rating	.55**	.55**		.49** (OWOS and 3, 4)

* Statistically significant at 5% level.
** Statistically significant at 1% level.

anonymously, but a check was made to make sure that an individual did not choose her own name. From these ratings a numerical score was determined for each individual for purposes of comparison with the individual's score on the Office Worker's Opinion Survey (OWOS).

Ratings by office managers, personnel managers, on-the-job supervisors, and the University coordinator. The same types of rating devices were used to obtain ratings for each individual office worker from each of the above persons.

A man-to-man rating device using a forced-choice technique was developed. This method required each rater to compare the personal traits of each office worker in his group with the personal traits of every other worker in the group. In this manner a numerical rating was obtained for each person and, also, it was possible to check the consistency of the rater.

A descriptive scale-type rating, employing five rather complete descriptions, was also used. This was not very helpful owing to the tendency of most raters to place all the workers in similar categories.

Discussion

The multiple-correlation coefficient between scores on the Survey and the three ratings was .51 (see Table 3), which is rather high for a validity coefficient.

The single validating criterion showing the highest correlation (.47) for the full-time group with scores on the Survey was the on-the-job supervisors' man to man rating. There may be several reasons for this high correlation, i.e., (*a*) supervisors may know the personal traits of office workers better than other raters, (*b*) the forced-choice technique used, and (*c*) supervisors may have a conception of what constitutes satisfactory personal traits for office work which is similar to that measured by the Office Workers Opinion Survey. At any rate it is of interest that the raters closely associated with the office workers gave ratings which correlated highest with scores on the Survey.

The low correlation between the descriptive scale ratings and the scores on the Survey was very likely due to the type of rating scale used. There were five descriptive categories in the scale and most raters were very reluctant to use the lowest two categories.

The various coefficients of correlation obtained for the group of 54 co-op students were very similar to those obtained for the full-time group.

The correlation between ratings by fellow office workers and scores on the Survey was rather low (.293) for the full-time group and rather high (.506) for the co-op group. This may have been due to the possibility that closer relationship existed among the co-op group than existed among the full-time group.

The data presented in the foregoing tables concerning the validity of the Office Workers Opinion Survey would seem to suggest that the Survey may be of use in helping companies evaluate the personal traits of office workers. The variation of OWOS scores among companies indicates that there is a wide variation of the personal traits of workers from company to company, but if each company would standardize the Office Work-

ers Opinion Survey on the basis of its own experience, the Survey could prove very helpful in their employment and promotion procedures.

Reliability. The method for estimating the reliability of an instrument using differential weighting of item responses developed by Hoyt and Stunkard [2] was used in determining the reliability of the Office Workers Opinion Survey. This method utilizes the following data: (*a*) the scores of each individual for all items, (*b*) the scores of each item for all individuals, (*c*) the sum of the scores on all items for all individuals, (*d*) the sum of the scores of all individuals for all items, (*e*) the number of items, and (*f*) the number of individuals.

A sample of one-half (64) of the unselected groups of 73 full-time office workers and 54 co-op students was obtained by coding each of the 127 papers and using a Table of Random Numbers. A reliability coefficient of .904 was obtained for the Survey, which is relatively high for an instrument of this type.

Summary

The purpose of this study was to investigate the feasibility of developing a technique which would aid in the differentiation of those employees who have desirable personal traits for office work from those who do not possess such personal traits.

The original instrument consisted of 457 opinion-type statements. These statements covered a variety of impressions concerning office work and there were five possible responses to each statement. The 457 statements were administered to two groups of "satisfactory" and "unsatisfactory" office workers chosen from 50 offices on the basis of their personal traits.

The item-selection process used to choose the items which differentiated between the two groups utilized: (*a*) the difference between the percentage of responses by the two groups to each category of each item, (*b*) the consistency of the direction of the differences between percentages in each item, (*c*) the significance of the differences between percentages, and (*d*) the statistic chi square. A total of 74 items were chosen as outstanding in their ability to discriminate between the two groups.

A scoring key was developed for these 74 items by assigning weights of plus one, zero, and minus one to the five responses of each item. This weighting was based on the direction and the significance of the differences between the percentages in each category. The 200 original surveys were scored on the basis of this scoring key and new criterion groups were established consisting of the upper and lower 25% of the original 200 surveys. A total of 150 items were chosen which discriminated between these two new criterion groups. The item-selection process used was similar to the one used to select the 74 items with the exception of chi square.

The 150 items, called the Office Workers Opinion Survey, were then administered to two unselected groups of office workers consisting of 73 full-time office workers and 54 business administration co-op students. Ratings concerning the personal traits of the individuals in these two groups were obtained from (*a*) fellow workers, (*b*) on-the-job supervisors, and (*c*) office managers or the University coordinator.

Multiple correlations obtained between scores on the Survey and the three ratings were .513 for the full-time group of office workers and .572 for the co-op students. These correlations were statistically significant at the 1% level. A reliability coefficient of .904 was obtained for the Survey using the analysis of variance method developed by Hoyt and Stunkard.

Conclusions. These data and statistics indicate that the Office Workers Opinion Survey shows promise for use in evaluating the personal traits of office workers. It should be emphasized, however, that any instrument of this type should be standardized by a particular firm on its own employees before individual results are used as a basis for making decisions concerning an applicant or employee.

Received June 21, 1955.

[2] Cyril J. Hoyt and Clayton L. Stunkard, Estimation of test reliability for unrestricted item scoring methods. *Educ. psychol. Measmt,* 1952, **12**, 756–758.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Internal Relations of Elemental Motions Within a Task

Norman B. Hall, Jr.

Dunlap & Associates, Inc.

Time-and-motion economy often has as its foundation an atomistic concept of human behavior. Revision and improvement of work cycles are made on the assumption that the individual elements which make up the total complex task can be removed or rearranged as though these elements are independent units. The elimination of a designated undesirable element is assumed to benefit the total time to perform the task. In authoritative texts can be found the following statements, "Lengthy hold or unavoidable delay operations offer particularly good possibilities for improving . . ." (**4**, p. 159), or, "The simo chart shows very distinctly where delays occur in the cycle and it aids in finding an effective way of eliminating the delays" (**1**, p. 101).

On the other hand this atomistic concept has been subject to criticism. "The individual working at a task, however simple, knits all the part activities into a whole; therefore the changing of any part activity may be expected to change the pattern of the larger task" (**3**, p. 268). The failure to consider the whole task and the fact that there exist internal relations between the elements is all too frequently overlooked by time-and-motion analysts, notwithstanding experimental results which demonstrate that removal of an element can result in an *increase* in the time of the remaining element of the cycle (**3**, p. 269).

As an illustration of how the times for work elements interact, Barnes and Mundel (**2**) reported the times for subjects to pick up and insert pins in holes varying from ¼ to 1 inch in diameter. The time in seconds to position a pin for insertion and to insert and remove a pin varied, as expected, with the size of the hole. There was also a variation, however, in the time to transport the pin from the supply to the hole. Such variation reflects an interaction between elements. It is the purpose of this study to provide further evidence concerning the extent of interaction of elements in an actual factory operation, thereby suggesting additional information upon which to reject the atomistic concept of human behavior in connection with micro-motion analysis.

Procedure

The industrial operation selected for study is the fastening of numbered or lettered plastic adding-machine key tops to metal key stems by means of acetate glue and a pneumatic press. The assembly process is as follows: The operator sits in front of a foot-operated pneumatic press which is mounted on a table. At her right hand is a bin containing "stems"; these are steel stampings, all of standard design. At her left is a bin containing the plastic key tops, all uniform as to shape and symbol at any given time. Each key top has a recess in the underside for the reception of the stem. The operator picks up a key top with her left hand, inspects it to orient the symbol, and places it upside down on the anvil of the press. The anvil is recessed to receive and hold the key top when pressure is applied. From the right-hand bin the operator picks up a stem, dips it momentarily in the acetate cup and positions it in the press, with the acetate-treated end over the key top. The operator then touches a foot pedal which opens a valve that admits air at 80 pounds per square inch pressure to an air cylinder and the press forces the stem into the recess in the key top. Upon releasing pressure on the foot pedal, the ram retracts and a small cam on the front of the press fixture throws the completed assembly into a chute leading to a bin underneath the table.

The machine operator chosen for this investigation was a highly skilled woman who had been employed on the operation for three years. Photographic samples of the key-top assembly process were made at eight intervals during the day—four in the morning and four in the afternoon. The times selected for these samples were: 7:45 A.M., 9:00 A.M., 10:00 A.M., 11:00 A.M., 12:30 P.M., 1:15 P.M., 2:30 P.M., and 3:30 P.M. These times were selected in order to obtain values for the elemental motions where the greatest time changes in the composite cycle might be expected, due to warm-up, practice, or the effects of continued activity. The duration of each sample was approximately 45 seconds. Slightly over 18 feet of film were used for each sample. Each sample yielded an average of 15 cycles. A frame-by-frame analysis of each sample was made according to the customary practice of micro-motion analysis (**4**), from which the simo-motion chart, Fig. 1, was constructed.

SIMO-MOTION CHART FOR KEY-TOP OPERATION

LEFT HAND DESCRIPTION	TIME IN MINUTES	SYMBOL	TIME SCALE	SYMBOL	TIME IN MINUTES	RIGHT HAND DESCRIPTION
CYCLE A						
POSITION KEY-TOP ON ANVIL FACE	.005	P		P	.002	POSITION STEM TO ACETATE CUP
TRANSPORT EMPTY to KEY-TOP BIN	.004	TE		TL	.007	TRANSPORT STEM TO KEY-TOP ON ANVIL
GRASP KEY-TOP	.003	G		P A	.005	POSITION STEM IN SLOT OF KEY-TOP START ASSEMBLY OF STEM AND TOP
BALANCING DELAY UNTIL the RIGHT HAND REACHES the POINT for SIMULTANEOUS WORK	.008	UD		P	.003	POSITION STEM AND KEY-TOP TO PRESS
				TE	.004	RETURN TO BIN
TRANSPORT OF KEY-TOP TO ANVIL INSPECTION(Visual) TO DETERMINE THE POSITION OF THE CHARACTER	.014	TL I		G	.003	GRASP OF STEM
				TL	.013	TRANSPORT STEM TO ACETATE CUP

FIG. 1. Simo-motion chart for key-top operation.

Results

The internal relations which exist between the individual therbligs (elementary subdivisions of a cycle of motions) are shown most clearly by a correlation value. These correlation values were computed by pairing the mean time for a therblig from each individual sample (based on 15 values) with the mean time of another therblig from the same sample. Forty such scores were available for the determination of each correlation value, as there were eight samples per day for five days. The number of possible pairings of therbligs for the right and left hand are 66 and the correlation values for each pairing are shown in Table 1.

A breakdown of these correlations is as follows:

Level of Significance	Negative	Positive
Greater than 5%	27	28
5%	4	3
1%	3	1
Total	34	32

Discussion

The highest positive correlation ($r = .57$) is between the therbligs "position to press" by the right hand and "unavoidable delay" by the left hand. A study of the simo-motion chart, Fig. 1, shows that maintenance of proper synchronism between the two hands requires the interaction of these two elements. The extent of this relation is given by the correlation value. An additional interaction between the left and right hand is provided by the significant correlation of $r = -.35$ between "grasp stem" (right hand) and "grasp key top" (left hand).

Statistical study of the values in the form of analysis of variance indicated that fluctuations in individual motions from sample to sample throughout the day are present and significant. The cycle time, however, remains constant. The constancy of production can be clearly observed in Table 2.

Since the cycle time remains constant, constant production is a concomitant fact. Internal compensation must therefore occur between the elemental motions of the cycle.

Correlation coefficients reflect the magnitude of this internal compensation and those which equal or exceed the 5% level of significance have been placed on Fig. 2. The most striking occurs in the negative correlation value of $-.93$ between the therblig "balancing delay until the right hand reaches the point for simultaneous work" and "grasp key top." This clearly shows that if time is not consumed in the "balancing delay" therblig, it is in the "grasp of the key top."

Table 1

Correlations Between Individual Therbligs
($N = 40$. See Fig. 1 for an explanation of the symbols used here.)

	GRASP-STEM	TL-STEM	POS. TO CUP	TL TO PRESS	POS. & ASS.	POS. TO PRESS	TE TO STEM BIN	GRASP-KEY TOP	UNV. DELAY	TL-KEY TOP	POS.-KEY TOP
GRASP-STEM											
TL-STEM	-.14										
POS. TO CUP	+.07	-.26									
TL TO PRESS	-.38†	+.33†	+.04								
POS. & ASS.	+.21	-.08	+.23	-.38†							
POS. TO PRESS	-.04	-.05	+.36†	+.36†	+.09						
TE TO STEM BIN	-.02	+.18	+.16	+.17	-.12	+.05					
GRASP-KEY TOP	-.35†	-.04	-.01	-.07	-.10	-.08	+.04				
UNV. DELAY	+.06	-.07	+.20	+.20	+.02	+.57*	-.06	-.93*			
TL-KEY TOP	+.16	+.43	-.03	+.22	-.05	-.06	+.05	-.06	-.33†		
POS.-KEY TOP	+.11	-.20	+.08	-.24	+.16	+.06	-.04	+.09	+.13	-.47*	
TE-KEY TOP BIN	-.07	+.08	-.03	+.13	+.01	-.12	+.26	-.08	-.10	-.01	-.42*

* SIGNIFICANT AT THE 1% LEVEL.
† SIGNIFICANT AT THE 5% LEVEL.

A further consideration of the values which are significant at the 5% level indicates that all correlations for the left hand are negative, while those of the right hand are both positive and negative. This is to be expected in the light of the following consideration: If all correlation values were positive, production would tend to vary with the therblig times. To have all values negative would be impossible. The combination permits synchronism and internal compensation. It is important to note that these correlations have meaning *only if* the cycle times are constant. Significant correlations like those in Table 1 could be obtained with no correlation between therbligs measured at any one time of day, but with significant differences between measurements of the same therbligs made at different times of day if the cycle time varied. In our case, however, such variation in time cycles during the day did not occur, as is reflected by Fig. 2.

Because the interrelation of the elemental motions of this and similar tasks is not adequately represented in the conventional form of simo-motion charts, it was considered valuable to construct a three-dimensional simo-motion chart (see Fig. 2). The elemental

motions of the right hand are placed on the upper circle, those of the left hand are on the lower circle. In both cases the sequence of elements is in a clockwise direction. The two circles are so placed that the corresponding movements of the hands are related as in the conventional simo-motion chart. Times for the elemental motions are represented in the three-dimensional simo-motion chart as proportional parts of the circle; total cycle time equals the full circle. Only correlation values which equal or exceed the 5% level of significance appear in Fig. 2. These correlation values demonstrate clearly the internal relation between each therblig.

It is very probable that if this three-dimensional procedure is used with other tasks, it may well throw further light on the synchronism and interaction of complex effort. Possibly in a more extensive study, the changes in these correlations from one time of day to another would reflect increasing compensatory synchronism related to fatigue. The absolute size of the average intercorrelation might well serve to indicate the degree of internal compensation for typical over-all variations in the therbligs.

Summary and Conclusions

Micro-motion samples of an industrial operation were taken eight times during the work day, Monday through Friday. These samples were analyzed by the customary micro-motion analysis technique. Each sample provided approximately 15 cycles (i.e., an individual element would occur 15 times in the cycle), and the average value for each individual element within a sample was determined. For the purpose of obtaining a correlation value, average times of individual elements within sample were paired. Forty such values were obtained and the relation of each elemental motion to every other was made, 66 values in all. Eleven of the correlations thus obtained equaled or exceeded the 5% level of significance.

The internal relations which exist between the elements of a manual operation are shown by these correlations between the time values

Table 2

Production Record for Key-Top Assembly Operation

(Average pieces produced per minute for half-hour intervals throughout the working day for each day of the week.)

Time	Monday	Tuesday	Wednesday	Thursday	Friday	Average	M.V.
7:30–8:00	22.7	19.0	20.3	21.2	21.1	20.9	.97
8:00–8:30	23.9	26.6	19.3	25.1	22.2	23.4	2.14
8:30–9:00	23.8	23.9	19.4	23.6	23.4	22.8	1.37
9:00–9:15	23.9	24.8	21.1	24.1	25.1	23.8	1.1
A.M. Rest Period							
9:25– 9:30	20.5	21.8	21.1	23.3	23.0	21.9	.96
9:30–10:00	21.5	23.0	20.8	23.3	22.2	22.2	.81
10:00–10:30	24.1	22.4	21.6	20.9	22.4	22.3	.82
10:30–11:00	23.8	24.7	21.5	24.2	23.4	23.5	.86
11:00–11:24	24.5	23.0	23.0	21.1	25.0	23.3	1.14
Lunch							
11:54–12:30	22.6	22.7	21.2	20.3	21.7	21.7	.76
12:30– 1:00	21.7	21.8	21.7	21.7	20.9	21.6	.26
1:00– 1:30	26.6	23.4	21.7	23.0	24.2	23.8	1.30
1:30– 1:45	22.6	24.3	23.9	22.4	22.4	23.1	.78
P.M. Rest Period							
1:55–2:00	21.6	23.7	21.9	20.9	20.0	21.6	.94
2:00–2:30	21.6	23.4	21.9	22.8	22.7	22.5	.58
2:30–3:00	22.2	21.0	23.5	22.0	20.7	21.9	.82
3:00–3:30	20.5	22.8	22.9	21.8	20.3	21.7	1.00
3:30–4:00	20.5	23.9	22.7	21.8	20.3	21.8	1.16

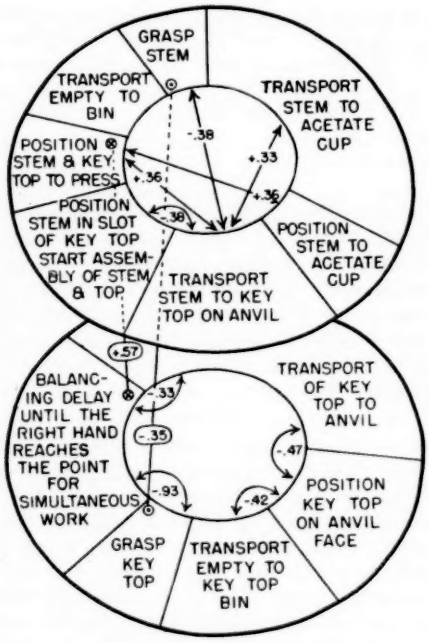

FIG. 2. Three dimensional simo-motion chart.

Therbligs for the right hand are on the upper circle, therbligs for the left hand are on the lower circle. The sequences of motions of each hand are read in a clockwise direction. The numbers shown are correlations; all equal or exceed the 5% level of significance.

for pairs of therbligs. If production is to remain constant, variations in time of the individual elements must be and are compensated for by either an increase or decrease in the time of the other elements which compose the cycle. This compensation may occur among the elements of one hand or between elements of both hands.

A three-dimensional simo-motion chart was used as a new method to represent the internal relations and to demonstrate that the industrial operation functions as, and must be viewed as, an integrated whole and not as a simple summation of independent units.

The results of this experiment provide evidence that the elements are related. The removal of one because of its undesirability may well be reflected in other elements. An atomistic concept of human behavior appears to ignore too much when considering motion economy.

Received May 13, 1955.

References

1. Barnes, R. M. *Motion and time study.* (2nd Ed.) New York: Wiley, 1940.
2. Barnes, R. M., & Mundel, M. E. Studies of hand motions and rhythm appearing in factory work. *Univer. Ia. Stud. Engng,* 1938, No. 12 (New ser. No. 350).
3. Ghiselli, E. E., & Brown, C. W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948.
4. Lowry, S. M., Maynard, H. B., & Stegemerten, G. J. *Time and motion study and formulas for wage incentive.* (3rd Ed.) New York: McGraw-Hill, 1940.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Factorial Analysis of Complex Psychomotor Performance and Related Skills

Edwin A. Fleishman and Walter E. Hempel, Jr.

Skill Components Research Laboratory
Air Force Personnel and Training Research Center[1]

Relative to research in other aptitude areas, little is known about the organization of abilities in the aptitude area of perceptual-motor skill. This is understandable in view of the technical and administrative difficulties involved in constructing and assembling the batteries of apparatus tests required for such research.

Even in the wartime Air Force classification research program, which involved the most extensive program of psychomotor test development ever attempted (**6**), no more than six psychomotor tests were ever included together in any of the experimental test batteries subjected to factor-analysis study. These analyses and subsequent postwar Air Force analyses (**3**, **4**, **10**, **11**) have been directed primarily at investigations of other aptitude areas (e.g., spatial visualization, perception, memory, integration, reasoning). In those analyses in which more than one psychomotor test appeared, a factor common only to these psychomotor tests and distinct from the printed tests was consistently demonstrated and labeled "Psychomotor Coordination." There was no opportunity, however, to construct experimental factor-analysis batteries, including a wider range of psychomotor tests, aimed at (*a*) splitting up this general psychomotor factor into possible more basic components, (*b*) clarifying the status of certain secondary psychomotor factors found on occasion in these studies, and/or (*c*) discovering additional psychomotor factors.

A series of such experimental psychomotor test batteries is currently being constructed and subjected to factorial study in this laboratory (see, e.g., **1**, **2**, **5**). Concurrently with these studies, however, factor analyses of certain relevant correlational matrices which exist from previous research would appear to provide additional insight and possible leads about the organization of abilities in this aptitude area.

Problem

The present paper, which represents another extension of the Air Force studies, is based on data previously published through the USAF School of Aviation Medicine, and originally collected in 1947–48 in a cooperative Air Force-Navy research project at Pensacola, Florida. In this project a wide variety of Air Force printed and apparatus tests were assembled and administered to over 1,000 Navy pilot candidates at the Naval Air Station. The test procedures and validities, reliabilities, and intercorrelations obtained have been described in detail elsewhere (**7**, **8**, **9**). The present paper describes a factor analysis of the intercorrelations among certain variables selected from the correlation matrix published by Payne (**9**); the analysis included 23 test variables of which 16 were apparatus psychomotor tests and 7 were printed tests designed as possible substitutes for the apparatus tests. Among the apparatus tests selected were the four tests that have consistently defined the "Psychomotor Coordination" factor in previous Air Force analyses. Also included were tests which had occasionally represented a factor tentatively labeled "Psychomotor Precision." It was also felt that the analysis would throw additional light on the role of spatial orientation and visualization abilities in psychomotor performance. In addition, it was possible to include the criterion of pilot success as one of the variables, since the validities of all the tests also had been determined on this sample. The

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7703. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. The data of this study have been reported in AFPTRC Research Bulletin 54–12.

criterion represents graduation or elimination from pilot training for reasons of flying deficiency. The inclusion of the criterion in the analysis was aimed at identifying the factor structure of the criterion, especially with regard to psychomotor skills.

It should be pointed out that after the completion of the present analysis and factor interpretations, the writers learned of a completely independent series of analyses of some of these same data by Roff (12). The present analysis, however, differs from Roff's analyses in several ways. Roff used different combinations of variables, favored an oblique reference frame, and did not include the pilot criterion in his factor analyses. In spite of these procedural differences, and some differences in factor interpretations, the two analyses showed some agreement on certain of the common factors identified. In those cases where factors identified in the two studies appear identical or highly similar, this will be pointed out in the text. One additional difficulty in making such comparisons is the fact that Roff sometimes included multiple scores from the same test. This often allowed the appearance of "doublet"-type factors specific to individual tests. The present analysis was confined to single scores from each test.

The Test Variables

Brief descriptions of the 23 selected test variables follow. More detailed descriptions of the apparatus tests may be found elsewhere (6). Descriptions of the printed tests may be found in (9).

Apparatus Tests: (code numbers refer to Air Force designations [6])

1. *Plane Control* (*CM817B*). The attitude of a model airplane is varied irregularly in its roll, pitch, and yaw axes by a motor-driven cam system. The examinee attempts to keep the airplane in a straight-and-level attitude by making compensatory adjustments of stick and pedal controls.

2. *Multidimensional Pursuit* (*CM813E*). Pointers on each of four meters (representing airspeed, r.p.m., bank, and turn indicators) vary irregularly, continuously, and independently of one another. The examinee attempts to keep all the pointers in the center of their respective scales by compensatory adjustments of simulated stick, rudder, and throttle controls.

3. *Rate Control* (*CM825A*). A target line moves back and forth across a curved scale, with frequent changes in direction and rate of movement. The examinee attempts to keep a pointer in coincidence with this line by adjustive manipulations of a knob control.

4. *Rotary Pursuit* (*CP410B*). The examinee attempts to keep a prod stylus in contact with a small metallic target, set in a rapidly revolving phonograph-type disk.

5. *Two-Hand Coordination* (*CM101B*). The examinee attempts to keep a target follower on a small target disk as the target moves irregularly and at varying rates. Movement of the target follower in the desired direction is controlled by the simultaneous rotation of two lathe-type handles.

6. *Complex Coordination* (*CM701E*). Patterns of lights are presented whose positions are to be matched by appropriate adjustments of stick and rudder controls. A correct response is accomplished only when both the hands and feet have completed and maintained the appropriate adjustments, at which point a new pattern of lights to be matched is presented.

7. *Rudder Control* (*CM120C*). The examinee sits in a mock airplane cockpit, which he attempts to keep lined up steadily with one of three target lights, as they come on in front of him. His own weight throws the seat off balance unless he applies and maintains proper correction by means of foot pedals.

8. *Direction Control* (*CP650A*). Patterns of lights appear in the form of a cross (to indicate the "heading" of an airplane). The pattern may appear in different parts of the display panel and may represent a small or large cross. Depending on the direction of "heading," the examinee must use both hands concurrently in manipulating the proper combination of switches or switches and buttons.

9. *Single Dimension Pursuitmeter* (*CM801B6*). The examinee makes compensatory adjustments to keep a horizontal line in a null position as it deviates from center in irregular fashion. Adjustments are made by in-and-out movements of a control wheel, dampened pneumatically to simulate the latency of aircraft controls.

10. *Compensatory Balance* (*CM510A*). The examinee controls the attitude of a platform containing a tortuous alley maze so that a gravity-activated ball will roll through the correct pathways (at 10 choice points) and not into blind alleys. Forward and backward tilt is controlled manually, side-to-side tilt by foot pedals.

11. *Controls Orientation* (*CP638A*). The examinee is confronted with an upright panel containing a circular arrangement of 16 lamps and a horizontal panel containing a corresponding arrangement of 16 toggle switches. For each successive problem he is required to push a switch that is the same number of switches distant from a variable reference point on the switch panel that a key lamp is from a stationary reference point on the lamp panel. The problems are further complicated in that the key

lamp is never lighted. Instead, its position is always midway between two lighted lamps.

12. *Pursuit Confusion* (*CM702B*). The examinee attempts to keep a stylus on a variable-speed target as it moves through a diamond-shaped slot. The task is complicated by the fact that the entire target area is visible only by mirror vision.

13. *Drift Correction* (*paced*). The display consists of two concentric circles of eight signal lamps each. For each stimulus setting, one light in the inner circle (representing plane headings) and one in the outer circle (representing wind direction) are lighted. The examinee pushes a toggle switch to the left or right to indicate the appropriate correction in heading.

14. *Discrimination Reaction Time* (*CP611E2*). The examinee manipulates one of four toggle switches as quickly as possible in response to a series of visual stimulus patterns differing from one another with respect to the spatial arrangement of their component parts (e.g., position of a lighted red lamp relative to a lighted green lamp).

15. *Complex Multiple Reaction* (*CP512AX1*). Visual and auditory stimuli alternate with one another as cues for *either* manual or pedal responses, respectively. The particular color or configuration of the visual cue determine which of three levers should be manipulated. The relative pitch of the auditory cue determines which of two pedals should be depressed.

16. *Santa Ana Finger Dexterity* (*CM116A*). This test involves a peg board containing 48 pegs having square bottom pieces and round tops. The examinee removes each peg in turn, rotates it clockwise through 180°, and reinserts it in its square hole with his preferred hand.

Table 1

Rotated Factor Loadings *

Variable	Factors** I PC-I	II PC-II	III SR-I	IV SR-II	V In	VI RC	VII P	VIII MD	IX Vz	X Res	h^2
1. Plane Control	25	50	00	07	15	15	02	24	09	05	43
2. Multidimensional Pursuit	20	38	18	23	36	05	07	24	07	−01	47
3. Rate Control	24	31	00	11	03	58	09	01	11	−08	53
4. Rotary Pursuit	47	08	13	17	05	24	16	35	−06	−03	49
5. Two-Hand Coordination	46	36	13	09	34	17	−04	10	04	15	55
6. Complex Coordination	50	37	16	22	36	10	05	10	21	−13	67
7. Rudder Control	40	56	11	−10	05	19	03	12	−05	−02	55
8. Direction Control	15	17	39	27	02	−06	02	21	44	09	53
9. Single Dimension Pursuitmeter	08	27	−06	16	04	55	07	05	07	06	43
10. Compensatory Balance	32	08	08	08	16	39	11	31	−10	−02	42
11. Controls Orientation	05	18	46	14	20	00	18	24	36	13	54
12. Pursuit Confusion	48	06	19	19	05	17	06	04	11	15	38
13. Drift Correction	03	20	40	15	21	08	02	15	00	13	31
14. Discrimination Reaction Time	10	17	37	50	17	14	08	34	10	−06	61
15. Complex Multiple Reaction	10	18	07	41	06	−02	05	19	18	13	31
16. Santa Ana Finger Dexterity	11	03	−07	26	19	09	20	38	12	11	34
17. Coordination (printed)	14	−01	03	00	09	03	42	24	00	02	26
18. Discrimination Pursuit (printed)	07	01	03	23	21	01	55	08	14	−02	43
19. Signal Discrimination (printed)	01	05	35	52	34	−07	40	10	07	−01	69
20. Complex Movements (printed)	06	03	13	09	40	01	00	07	32	16	32
21. Coordinate Movements (printed)	10	09	36	18	44	00	19	02	11	06	43
22. Directional Control (printed)	10	−10	38	03	28	18	10	02	34	−11	41
23. Discrimination Reaction Time (printed)	11	12	23	41	25	−06	35	04	08	09	45
24. Criterion	64	23	22	02	−02	−04	07	08	07	−03	53
$\Sigma a^2/k$	08	06	06	06	05	04	04	04	03	01	

* Decimal points are omitted.

** Factors are identified as follows: I—Psychomotor Coordination I; II—Psychomotor Coordination II; III—Spatial Relations I; IV—Spatial Relations II; V—Integration; VI—Rate Control; VII—Perceptual Speed; VIII—Manual Dexterity; IX—Visualization; X—Residual.

Printed Tests: (code numbers are Navy designations [9])

17. *Coordination Test* (*NavMed 1215*). The pathways on a series of superimposed mazes must be traced accurately and rapidly.

18. *Discrimination Pursuit* (*NavMed 1216*). Twenty-five rows of 5 circles each are presented and these 5 circles are of varying diameters. Different problems are presented in which circles of certain diameters in certain columns are to be connected with certain other circles as rapidly and accurately as possible.

19. *Signal Discrimination* (*NavMed 1212*). Each of a series of items presents 4 circles. One circle contains a zero, another a plus mark, and the remaining two are empty. The task is to determine if the plus mark is right, left, above, below, or diagonal to the zero.

20. *Complex Movements* (*NavMed 1214*). Each problem presents a vertical line and above it a horizontal line. On each of these lines is a circle and a dot. Below the lines is a grid of 49 squares, one of which contains an "S." The examinee indicates in the grid, from the letter "S," the direction required to "move" the dots on both lines simultaneously into their respective circles.

21. *Coordinate Movements* (*NavMed 1211*). Each problem presents a series of arrows which vary in curvature and represent different directions of movement. The thickness of each arrow represents distance of movement. To the left of these arrows are a dot and a circle. The examinee selects the one symbol that will move the dot inside the circle.

22. *Directional Control* (*NavMed 1213*). Each problem presents nine boxes side by side, with a target circle above the middle box. Also pictured is an arrow whose head indicates direction and whose length indicates distance of movement. The examinee indicates from which box it must be launched to have its head reach the target.

23. *Printed Discrimination Reaction Time* (*CP634A*). Each item simulates a setting of the apparatus test of the same name (see Test 14 above), except responses are made by pencil in one of four slots on answer sheets. The slots are arranged in the same up-down, left-right pattern as are the toggle switches on the apparatus version.

The Criterion

24. Graduation versus elimination of Navy midshipmen from flying training, where eliminations were due to flying deficiency.

Results

Ten factors were extracted from the intercorrelations among these variables by the Thurstone centroid method (13).[2] The axes defined by these centroid factor loadings were then rotated orthogonally to simple structure and positive manifold, using Zimmerman's graphical procedure (14). Table 1 presents the complete matrix of rotated factor loadings.

Interpretation of the Factors

The rotated factors were interpreted for psychological meaningfulness. Loadings above .20 were considered significant. For presentation, only loadings above .30 for test variables and those above .20 for the pilot criterion are listed in turn for each factor.

Factor I is identified as Psychomotor Coordination—I.

Variable No.		Loading
6	Complex Coordination	.50
12	Pursuit Confusion	.48
4	Rotary Pursuit	.47
5	Two-Hand Coordination	.46
7	Rudder Control	.40
10	Compensatory Balance	.32
24	Criterion	.64

Previous Air Force studies have consistently identified a factor common to four of the psychomotor tests which have been operational (Variables 4, 5, 6, 7). This factor was called Psychomotor Coordination and defined broadly as representing the coordination of the larger muscles of the body in movements of moderate scope. The present battery represents the first time the four tests defining this factor have been administered in combination with a large number of other psychomotor tests. As a consequence, it appears that this Psychomotor Coordination factor has been split into several other factors. Another possibility is that we have simply uncovered additional psychomotor factors present in these tests. The distinguishing characteristic of the present factor appears to be that fine, sensitive, highly-controlled adjustments are required in movements quite restricted in scope. For example, in certain tasks movement of controls into approxi-

[2] The correlation matrix and the matrix of centroid factor loadings have been deposited with the American Documentation Institute. Order Document No. 4726 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

mately the proper position is first necessary, but then more delicate adjustment of the controls may be necessary to locate the exact target position and to maintain it. Such very fine controlled muscular adjustments seem crucial at some stage of performance in all the tasks loaded on this factor. For the present, then, this factor is interpreted as representing the ability to control muscular movements involved in making fine, accurate adjustments. It can also be seen that this factor contributes a large proportion of the variance in the pilot criterion.

Factor II is identified as Psychomotor Coordination—**II.**

Variable No.		Loading
7	Rudder Control	.56
1	Plane Control	.50
2	Multidimensional Pursuit	.38
6	Complex Coordination	.37
5	Two-Hand Coordination	.36
3	Rate Control	.31
24	Criterion	.23

This factor appears to involve coordination between muscle groups in making more gross adjustments, where the use of more than one body member is required. The tests most heavily loaded on this factor all emphasize these more extensive coordinate adjustments. In fact, the first four tests on this factor involve coordinate leg movements. Three of these four also involve arm movements, and the fifth test involves coordinate movements of both arms only. It is true that several of these tests involve both Psychomotor Coordination—I (fine adjustment) and Psychomotor Coordination—II. However, the present distinction between these two factors rests primarily on the presence of Rotary Pursuit and Pursuit Confusion in Psychomotor Coordination—I and their absence from the present factor. In both these tests skill around the target area is the important feature, no adjustment is required from one target to another, and only one body member is involved. It can also be seen that the Psychomotor Coordination—II factor contributes in some degree to the pilot criterion.

Factor III is identified as Spatial Relations I (stimulus interpretation). The precise definition of the factor of Spatial Relations has been the subject of considerable study (e.g., **3, 4, 15**). It has been defined broadly by Guilford (**3**) as the ability to relate different stimuli to different responses, where either stimuli or responses are arranged in spatial order. However, it has not been clear whether the appreciation of spatial arrangement of stimuli or of responses separately is the key to the factor. Several more specific hypotheses have been proposed (e.g., **3, 15**). According to one hypothesis, it is the ability to make discriminations as to directions of motion (e.g., up, down, left, and right). Other hypotheses suggest the essence of the space factor could be (*a*) ability to perceive visual-spatial relationships, (*b*) ability to organize movements in spatial-determined order, or (*c*) ability to relate specific spatial locus or arrangement within the stimulus pattern with specific locus or arrangement within the response pattern.

The present analysis contributes certain leads toward resolving these questions. The presence of a wider variety of printed and psychomotor tasks, with obvious spatial characteristics, has apparently allowed a separation into at least two spatial factors. Each of these factors is consistent with more recent emphases (**4, 15**) that spatial relations involves orientation with respect to one's own body.

The first of these factors is defined by the following variables:

Variable No.		Loading
11	Controls Orientation	.46
13	Drift Correction	.40
8	Direction Control	.39
22	Directional Control (printed)	.38
14	Discrimination Reaction Time	.37
21	Coordinate Movements (printed)	.36
19	Signal Discrimination (printed)	.35
24	Criterion	.22

The tests most heavily saturated with this factor involve the interpretation of the spatial relations of the *stimuli* before the proper response can be determined. Moreover, loadings appear to be largest in the tests presenting the most complex stimulus patterns, in which interpretation of the proper spatial relationships of the components of the pattern are the most difficult parts of the task. Orientation, however, is still with respect to one's own body, since in all the tasks there is a

logical relationship between the responses required and the spatial characteristics of the stimulus. The emphasis, however, appears to be on the ability to perceive and interpret these visual-spatial characteristics of the stimulus. Roff (12) identified a similar factor in which the element of skill in responding was apparently subordinate to the skill in differentiating the stimulus. Roff, however, does not refer to the spatial characteristics of the task and prefers to call the factor simply Visuo-Motor Discrimination.

This is the spatial factor which contributes to the pilot criterion.

Factor IV is identified as Spatial Relations II (response orientation or choice).

Variable No.		Loading
19	Signal Discrimination (printed)	.52
14	Discrimination Reaction Time	.50
23	Discrimination Reaction Time (printed)	.41
15	Complex Multiple Reaction	.41

It is involved in tasks which have a number of possible responses which are spatially distinct. The important distinction between this factor and Spatial Relations I is that the patterns of stimuli need *not* be spatially arranged and need not be difficult to discriminate. The crucial parts of the tasks appear to be in decision and choice of response. This factor appears to represent the abilities to make rapid discriminations as to directions of motion. Even the printed tests on this factor involve the *making* of an appropriate response (e.g., checking slots arranged in an up, down, right, left pattern, or drawing a series of lines in the correct directions). The Complex Multiple Reaction Test is the purest measure of this factor, and here the stimuli are not spatially arranged (discriminations are to be made primarily between colors of light and/or between visual versus auditory cues). The responses required are quite complex, however (one of three levers, or right or left foot pedal). Even the Finger Dexterity Test (with a loading of .26) requires a decision as to counterclockwise or clockwise movement of pegs.[3]

The present analysis does not allow a crucial test of the distinction between these two spatial factors since many of the tasks involve spatial arrangements of both stimuli and responses, but the present results do suggest the kind of experimentation required to substantiate or reject the present hypothesis. This would require an analysis of a battery of tests, some of which involve simple nonspatial responses and complex spatial stimulus characteristics, while others require complex choices of responses arranged spatially, but involve no spatial interpretations of stimuli (e.g., The Complex Multiple Reaction Test).

Factor V is tentatively identified as *Integration*. It appears in both printed and apparatus tests.

Variable No.		Loading
21	Coordinate Movements (printed)	.44
20	Complex Movements (printed)	.40
2	Multidimensional Pursuit	.36
6	Complex Coordination	.36
5	Two-Hand Coordination	.34
19	Signal Discrimination (printed)	.34

The apparatus tests with highest loadings on this factor all require simultaneous responses of more than one body member. Successful performance in each of these tasks depends on the proper coordination or merging of movements of both hands or hands and feet. For example, in the Two-Hand Coordination Test, movements of one hand alone move the target follower to the right or to the left, while movements of the other hand alone control the target follower toward or away from the subject. Proper combinations of movements together, however, can move the target follower in any *resultant* direction. Similar requirements are imposed in the other apparatus tasks.

The presence of certain printed tests on this factor suggest further hypotheses about its nature. For example, in the Coordinate Movements Test, one type of arrow represents left-right movement, another type represents up-down movement, and still another type represents no movement. Thickness of arrow represents distance of movement. The examinee must select from five combinations

[3] Possible disturbing elements in our discussion of these Spatial Relations Factors are the loadings of the Complex Coordination Test, which have generally appeared on a Spatial Relations factor. This test is not loaded on Spatial Relations I and has a loading on only .22 on Spatial Relations II.

of these symbols the one symbol which will give the proper *resultant* response (e.g., the arrow which will "move" a dot exactly inside a circle located in a certain direction and at a certain distance from it). In the Complex Movements Test, the examinee must choose the *one* response which will produce the proper extent and direction of "movement" in *two* separate dots. Any of the wrong choices will leave one or both of the dots short of the target circle. In either of these two printed tests the examinee must utilize and apply the information provided by a number of separate symbols into an integrated single response.

It therefore appears that this factor involves the ability to utilize and coordinate a number of disparate cues and activities quickly and accurately in order to produce an appropriate integrated response.[4] Roff (12) also identified a factor containing these same apparatus tests, which he called *Integrated Coordination*. However, these printed tests did not appear on this factor in his analysis. It appears that Roff's factor is the same as our Psychomotor Coordination II and that our Integration factor was not identified in his analysis.

Factor VI is defined as *Rate Control.*

Variable No.		Loading
3	Rate Control	.58
9	Single-Dimension Pursuitmeter	.55
10	Compensatory Balance	.39

This factor is common only to certain of the apparatus tests, and to none of the printed tests. It appears to involve an aspect of pursuit. Each of the tasks requires the examinee to make anticipatory adjustments relative to changes in speed and direction of a continuously moving object. For example, in the Rate Control Test, the examinee must keep a pointer on a target line as the line moves across a scale with frequent changes in rate of acceleration and direction. In the Single Dimension Pursuit task, a horizontal line must be kept in a null position as it deviates in an unpredictable manner as to direction and rate of movement. The nature of the Compensatory Balance Test may seem ostensibly to differ from the other tests on this factor. It can be seen, however, that judging and compensating for the rate of movement of the steel ball as it approaches the various choice points in the maze pattern is crucial in keeping the ball in the correct pathway. An interesting feature of this factor is that it cuts across the traditional categorization of pursuit tasks into "following pursuit" (e.g., Rate Control),[5] and "compensatory pursuit" (e.g., Single Dimension Pursuit) tasks. These results suggest that from the point of view of individual differences, this distinction may be arbitrary, and the nature of such tasks may better be thought of in terms of a third underlying variable. This is confirmed by Roff's analysis (12) which found a factor defined only by the Rate Control and Single Dimension Pursuitmeter tests. He defined his factor as *Eccentric Pursuit,* but the present analysis suggests the nature of the factor may be broader than this.

Factor VII is identified as the well-established factor of *Perceptual Speed.*

Variable No.		Loading
18	Discrimination Pursuit (printed)	.55
17	Coordination (printed)	.42
19	Signal Discrimination (printed)	.40
23	Discrimination Reaction Time (printed)	.35

This factor involves facility in making rapid comparisons of visual forms and the notation of similarities and differences in form and detail. The nature of the tests with highest loadings are quite similar to those previously identified with this factor (e.g., picking out certain forms presented with a number of irrelevant forms, tracing maze lines, etc.). This factor does not appear in any of the present apparatus tests. This factor is practically identical to one identified in Roff's analysis (12).

Factor VIII is identified as *Manual Dexterity.*

[4] The relationship between this factor and the Integration factors identified in the Air Force analyses (3) is in doubt. The Air Force analyses failed to find an Integration factor in the psychomotor tests. However, the present integration factor seems of a different type from those identified previously.

[5] Rotary Pursuit, another "following pursuit" task, has a loading of .24 on this factor and differs from the other tasks in that the course and rate of its target is more predictable.

Variable No.		Loading
16	Santa Ana Finger Dexterity	.38
4	Rotary Pursuit	.35
14	Discrimination Reaction Time	.34
10	Compensatory Balance	.31

This factor appears to be the same one identified in a few of the previous Air Force analyses as *Psychomotor Precision* and tentatively defined as the ability to make precise manipulations under speed conditions. However, this definition appears much too broad. Other definitions have identified this factor with finger dexterity. The present interpretation of this factor as Manual Dexterity appears much less strained, especially in view of the significant loadings of Rotary Pursuit and Compensatory Balance, which do not emphasize skillful finger movements but do involve skillful arm-hand movements. Similarly, the Discrimination Reaction Time Test may be viewed as involving skillful, well-directed arm-hand movements. Moreover, in our previous studies involving dexterity tests (1, 2) the Santa Ana Test was found to contain a high loading on a factor emphasizing more gross arm and hand movements (Manual Dexterity) and a low loading on a factor identified as emphasizing finger coordination (Finger Dexterity). Manual Dexterity is also distinguished from Psychomotor Coordination —II in that the latter involves more coordination between muscle groups, is not entirely restricted to arm-hand movements, and does not appear as concerned with speed. That this factor does not contribute to the pilot criterion has confirmed earlier findings (3). The factor has been found important in the bombardier criterion, however.

Factor IX is identified as *Visualization*.

Variable No.		Loading
8	Direction Control	.44
11	Controls Orientation	.36
22	Directional Control (printed)	.34
20	Complex Movements (printed)	.32

It has been defined as the ability to make mental manipulations of visual images. In established tests of the Visualization factor, it is usually necessary to move, turn, twist, or rotate the stimulus situation in imagination and to recognize a new appearance, position, or condition after prescribed manipulations. A separation between visualization and spatial-relations factors occurred first in certain Air Force analyses (3). The two apparatus tests on this factor also have strong loadings on Spatial Relations I. Thus, interpretation of the spatial components of the complex stimulus patterns presented in relation to the responses required have been hypothesized to account in some measure for performance on these tests. It seems reasonable from the nature of these tasks, however, that the ability to manipulate or rotate the entire stimulus pattern mentally into a new position also is necessary for effective performance on these particular, more difficult tasks.[6] It should be pointed out that none of the better established tests of Visualization have been included in the present analysis. However, each of the four tests loaded on the present factor exhibit higher correlations with such established tests of visualization (e.g., Pattern Comprehension, Spatial Visualization) than do tests not loaded on this factor (see 9).

Factor X is a residual factor containing only insignificant loadings.

Conclusions

1. Of the nine factors identified, four were confined to the apparatus psychomotor tests.

a. Two psychomotor factors were identified among the tests which had previously identified a single factor called Psychomotor Coordination in many previous Air Force analyses. One of these factors appears to emphasize fine, highly controlled sensitive adjustments in movements quite restricted in scope.

b. The second factor seems to involve more gross coordination in making more extensive movements and was best measured by tests requiring the use of several body members simultaneously.

c. The status of the factor previously called Psychomotor Precision in the Air Force research was clarified somewhat. This factor appears more adequately described as Manual Dexterity.

d. Another psychomotor factor, called Rate Control, was restricted to certain apparatus tests which required the ability to make con-

[6] This finding supports the recent contention by Zimmerman (15) that Spatial Relations and Visualization may be on some kind of difficulty continuum.

tinuous motor adjustments relative to changes in speed of direction of a continuously moving object. This factor was found to cut across the traditional classification of "compensatory" versus "following" types of pursuit tasks.

2. Four factors were found measurable by both apparatus tests and certain printed tests.

a. An Integration factor was found in tests which required the ability to utilize and apply a number of disparate cues and activities quickly into an integrated resultant response. Its relation to integration factors found in previous research needs to be determined.

b. Two separate Spatial Relations factors were isolated which suggest leads for resolving certain questions in this area. One of these factors emphasizes the ability to interpret spatial characteristics of the stimulus situation, while the other is more identified with directional discrimination and orientation of movement patterns.

c. The separation of a Visualization factor from Spatial Relations has been confirmed. Of interest in this connection is the presence of Visualization in certain of the perceptually more difficult apparatus tests.

3. A Perceptual Speed factor was found confined to printed tests.

4. Over 50 per cent of the variance in the criterion of pilot proficiency was accounted for by the factors identified. Indications are that better tests are especially needed for the most valid factor, Psychomotor Coordination—I. The other two factors contributing significantly were Psychomotor Coordination—II and Spatial Relations I.

5. Contrary to previous belief that motor skills are narrow in scope and highly specific to the task, the present results confirm that there are certain broad group factors of psychomotor skill which may account for performance on a wide variety of different psychomotor tasks.

6. In general, some of the factors defined in psychomotor tests may also be sampled by printed tests. Although the distinction between "motor" and "nonmotor" factors may be considered somewhat arbitrary, there is evidence that psychomotor tests sample some factors not covered by any other kinds of tests.

Received June 15, 1955.

References

1. Fleishman, E. A. Dimensional analysis of psychomotor abilities. *J. exp. Psychol.*, 1954, **48**, 437–454. (A factorial study of psychomotor abilities. *USAF, Personnel & Training Res. Cent., Res. Bull.*, 1954, No. 54–15 [AFPTRC–TR–54–15].)
2. Fleishman, E. A., & Hempel, W. E., Jr. A factor analysis of dexterity tests. *Personnel Psychol.*, 1954, **7**, 15–32.
3. Guilford, J. P. (Ed.) *Printed classification tests.* Washington: U. S. Government Printing Office, 1947. (*AAF Psychol. Program Res. Rep.* No. 5.)
4. Guilford, J. P., Fruchter, B., & Zimmerman, W. S. Factor analysis of the Army Air Forces Shepard Field Battery of experimental aptitude tests. *Psychometrika*, 1952, **17**, 45–68.
5. Hempel, W. E., Jr., & Fleishman, E. A. A factor analysis of physical proficiency and manipulative skill. *J. appl. Psychol.*, 1955, **39**, 12–16. (*USAF, Personnel & Training Res. Cent., Res. Bull.*, 1954, No. 54–34 [AFPTRC–TR–54–34].)
6. Melton, A. W. (Ed.) *Apparatus tests.* Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
7. Page, H. E. (Ed.) The pilot candidate school research program: historical background and organization. *USAF Sch. Aviat. Med., Joint AF-Bu Med Res. Project Rep.* No. 1, 1951.
8. Payne, R. B. (Ed.) The pilot candidate selection research program: implementation procedures. *USAF Sch. Aviat. Med., Joint AF-Bu Med Project Rep.* No. 2, 1950.
9. Payne, R. B. (Ed.) The pilot candidate selection research program, IV: Test validities and intercorrelation. *USAF Sch. Aviat. Med., Joint AF-Bu Med Project Rep.* No. 4, 1952.
10. Roff, M. F. A factorial study of tests in the perceptual area. *Psychometric Monogr.*, 1952, No. 8.
11. Roff, M. F. Personnel and classification procedures: spatial tests. *USAF Sch. Aviat. Med. Project Rep.*, 1951.
12. Roff, M. F. (Ed.) The pilot candidate selection research program, V: A factorial study of the motor aptitudes area. *USAF Sch. Aviat. Med., Joint AF-Bu Med Res. Project Rep.* No. 5, 1953.
13. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.
14. Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–55.
15. Zimmerman, W. S. A revised orthogonal rotational solution for Thurstone's original primary abilities test battery. *Psychometrika*, 1953, **18**, 77–93.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Spatial Factors in Check Reading of Dial Groups [1]

Robert S. Lincoln [2] and Emanuel Averbach

The Johns Hopkins University

It would seem reasonable to expect that reading habits will be influential in determining the pattern of scanning that an observer might use in viewing visual displays. If displays are scanned in a manner similar to that used in reading, there should be some relationship between the accuracy with which elements of the display are read and the spatial location of those elements. This idea is in line with the results of an experiment by White, Warrick, and Grether (2) who found that observers tended to fixate initially on the upper half of a panel of check dials while they were looking for misaligned pointers in the panel. Furthermore, their observers missed 22% more deviant pointers in the bottom half of the panel than in the top half. The present experiment was designed to test more completely the effects of spatial location on the accuracy with which check dials are read as well as the relation between this accuracy and reading habits. The design also made it possible to test the effects of different durations of exposure of the dial displays.

Method

Apparatus. Figure 1 is a schematic diagram of one of the panels used in this experiment. On the actual panels the dials were 1.6 in. in diameter, and were separated from each other by a space of .2 of an inch. For purposes of analysis the panels were divided into four quadrants with three positions in each quadrant. In Fig. 1 the dotted lines separate the four quadrants of the panel, and the letters within the dials indicate the three quadrant positions, "C" indicating corner, "S" indicating side, and "M" indicating a middle position with respect to the entire panel. No distinction was made between the two side positions in each quadrant. They were treated as one position. The letters, words, and dotted lines did not appear on the test panels. The observer's (*O*'s) task was to detect the pointers that deviated from the null position. The *O*s indicated their detections by putting check marks in the appropriate spaces on answer sheets that resembled the dial panels. The pointers that were misaligned always deviated 15 degrees in a clockwise direction from the null position. There was a 7-degree separation between the deviant pointers and the upper edge of the normal range indicated on each dial face.

A total of 42 panels was used. Four deviant pointers appeared on each panel. Panels were divided into 14 sets of three panels each. In all sets a deviant pointer appeared once in each of the three positions within each of the quadrants, making a total of 12 deviant pointers in each set of three panels.[3] Combinations of deviant pointers within a panel were chosen in a random manner. Since there were two side positions in each quadrant, deviant pointers were randomly assigned between them. An effort was made to ensure that on most of the individual panels there were deviant pointers in each quadrant and in each of the three quadrant positions. Figure 1, for example, shows deviant pointers in: (*a*) top-left quadrant, corner position; (*b*)

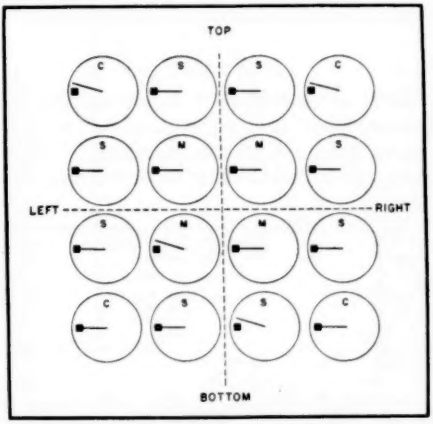

FIG. 1. Schematic diagram of a dial panel. The dotted lines separate the quadrants of the panel. The letters indicate positions within quadrants. The letter "C" indicates a corner position, "S" indicates a side position, and "M" indicates a middle position.

[1] This experiment was done under Contract N5–ori–166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166–I–197, Project Designation No. NR 145–089, under that contract.

[2] Now with the Systems Development Division, The RAND Corporation, Santa Monica, California.

[3] After the experiment had been run, an error was discovered in the construction of one panel in Set 11. In this set, one middle position was omitted, and a side position was represented in its place. The effect of this error was negligible.

top-right quadrant, corner position; (*c*) bottom-left quadrant, middle position; and (*d*) bottom-right quadrant, side position.

The panels were displayed in an electronic tachistoscope that has been previously described (1). The panels were viewed at a distance of 23 in. from the *O*'s eyes. The brightness of the panels was approximately 2.7 millilamberts. Light adaptation was maintained between exposures with a special adapting light.

Procedures. Three different groups of 14 undergraduate students each were used in the experiment. For one group the panels were exposed for a period of .35 sec. The exposure for the second group was .70 sec., and for the third group was 1.40 sec. The 14 sets of panels were presented to the *O*s in orders determined by a 14×14 latin square. One square was used for all three groups. The same 14 sets of panels were presented twice to each *O*. A 5-min. rest break was given after the first presentation of the panels.

Before the test panels were presented, all *O*s received practice in identifying various symbols placed in different positions on the practice panels. They were also shown drawings that indicated the null position and the appearance of a deviating pointer. The *O*s were not told the number of deviant pointers that would appear on the panels. They were instructed not to guess, but were encouraged to try to indicate the positions of all of the deviating pointers that they actually detected. The data from two *O*s were discarded because they very obviously failed to follow the instructions concerning guessing.

Results

Most of the analysis of results concerns the percentages of deviant pointers that were detected under the various experimental conditions. It sometimes happened, however, that *O*s would indicate a deviant pointer in a location where the pointer did not deviate. These errors of localization will also be described.

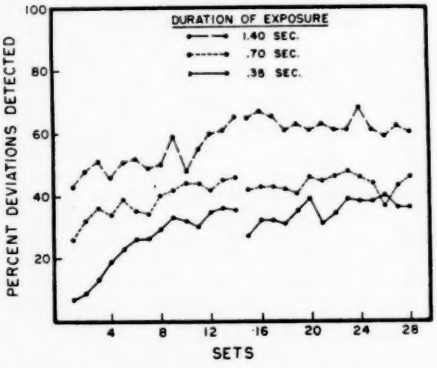

FIG. 2. Practice curves for different durations of exposure.

All statements of statistical significance are based upon nonparametric tests. When a comparison involved two different groups of *O*s, a test for unpaired replicates was used. When a comparison involved the same group of *O*s under a number of conditions, Friedman's chi-square test was used. Both of these tests have been described by Wilcoxon (3).

Effects of practice and duration of exposure. The curves shown in Fig. 2 reflect both the effects of practice and duration of exposure. The breaks in the curves represent the rest pause that separated the two presentations of the same 14 sets of panels. These *O*s improved their accuracy considerably, but most of the improvement took place in the first half of the experimental period.

The effects of exposure duration are very consistent and highly significant ($p < .01$). Since the durations are themselves geometrically related, the relation between percentage of detections and exposure duration is nonlinear during the first half of the practice period. This relationship appears to become more nearly linear during the second half of the practice period.

Despite the obvious increase in the percentage of detections that occurred with longer exposures, *O*s made significantly more errors of localization as exposure time increased ($p < .01$). The longer the exposure, the more frequently the *O*s indicated deviant pointers in dials where deviations did not exist. This effect may have occurred because the *O*s saw more deviant pointers at the longer exposures, but were unable to remember exactly where they had seen them. With the shorter exposures they saw fewer deviant pointers and, therefore, may have had less trouble remembering where they had seen them. At the longest exposure 9.7% of the reports of deviant pointers involved an error of localization. Since the *O*s were instructed not to guess, it is likely that the curves in Fig. 2 show the percentages detected and remembered. The curves apparently underesti-

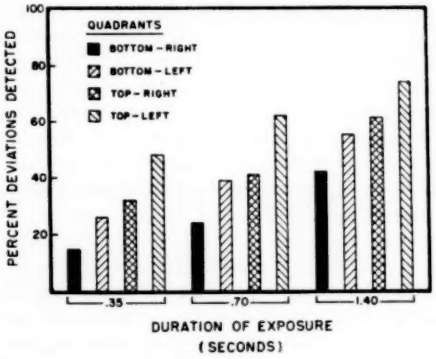

FIG. 3. Percentage of detections in various quadrants for different exposures.

mate the number of deviant pointers actually detected.

Quadrant effects. If *O*s did scan the dial panels as though they were reading them, the smallest number of detections would be expected in the bottom-right quadrant. Progressively more detections would be expected in the bottom-left and top-right quadrants. The greatest number of detections would be expected in the top-left quadrant.

Figure 3 shows percentage of detections by quadrants. For each duration of exposure the percentage of detections increases in the expected manner. The differences among these quadrants are significant ($p < .01$). Figure 3 also indicates that the effects of exposure are very consistent for each quadrant considered separately.

Effects of position. Within each quadrant separate dials were located either on a side, in a corner, or in a middle position on the panel. The experimental panels were designed so that deviant pointers appeared in either one or the other side positions as often as in the corner or middle positions. Because of this arrangement it is possible to compare the number of detections made in each of these three positions within and between quadrants. Of particular interest are the practice curves obtained for each position considered separately. As Fig. 2 indicated, performance improved considerably in the first half of the practice period. In that figure, however, the different positions were pooled. It would have been possible to obtain the curves of Fig. 2 if improvement had been shown in detecting deviant pointers in only one position while performance remained stable for the other positions. Actually improvement was observed for each of the three positions. This result is pictured in Fig. 4 in which exposures and quadrants are pooled. The differences among positions are significant ($p < .01$). The data from only the first half of the experiment are pictured in Fig. 4 because most of the practice effect appeared in the early sets. The relative positions of the curves remained the same in the second half of the experiment.

No fixation point was provided on the dial panels used in this experiment. As Fig. 4 shows, however, the greatest percentage of detections was made in the middle positions on the panels from the very beginning of the experiment. This result suggests that these *O*s immediately established their own initial fixation point in the center of each panel. Subsequent fixations on a panel probably fol-

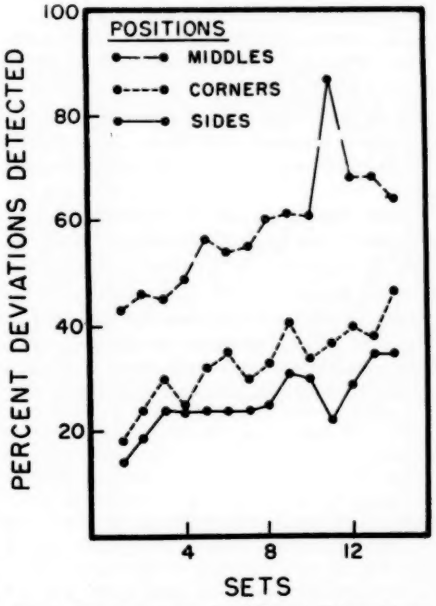

FIG. 4. Practice curves for different dial positions.

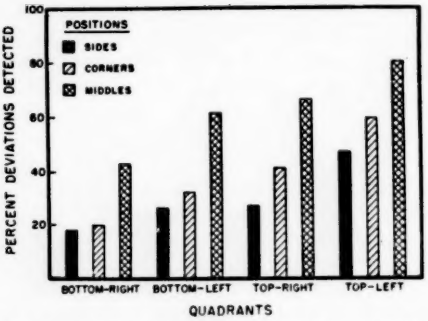

FIG. 5. Percentage of detections for different positions within quadrants.

lowed according to established reading habits, as the bars in Fig. 3 indicated.

In Fig. 4 the aberrant points in Set 11 for the middle and side positions seem to have resulted from the error in panel make-up that was previously described. One panel contained an extra side position and was short one middle position; *O*s detected about the same absolute number of deviant pointers as they detected in Sets 10 and 12. Apparently these *O*s were detecting as many of the deviant pointers as they could, and a few more or less in one position did not make much difference. Therefore, when the number of detections is plotted in Fig. 4 as a percentage of possible detections, the percentage of actual detections becomes abnormally high for the middle positions and low for the side positions.

Since the three positions were equally represented in each of the quadrants, it is possible to compare the frequency of detection for each position within each quadrant. The results of this comparison are pictured in Fig. 5 in which exposures are pooled. It is apparent that the positional effects are highly consistent within the various quadrants. In addition the percentage detected for each position changes quite consistently between quadrants. Deviant pointers were detected about 86% of the time when located in the middle position of the top-left quadrant with an exposure of 1.40 sec. They were detected about 5% of the time when located in a side position of the bottom-right quadrant with an exposure of .35 sec.

Summary

Observers were required to detect deviant pointers within a display panel of 16 circular dials. For each dial the null point was located in the 9 o'clock position. Throughout the experiment the spatial locations of the deviant pointers within a panel were controlled in such a way that it was possible to determine the percentage of deviations detected as a function of quadrant location and position within quadrants. The consistency of these spatial effects was determined over three durations of panel exposure.

The results showed that spatial location was an important determinant of the number of detections that were made. The pattern of detections that appeared seems to confirm the idea that the scanning habits which observers use are highly related to previously learned reading habits.

Received May 31, 1955.

References

1. Merryman, J. G., & Allen, H. E. An improved electronic tachistoscope. *Amer. J. Psychol.*, 1953, **66**, 110–114.
2. White, W. J., Warrick, M. J., & Grether, W. F. Instrument reading. III: Check reading of instrument groups. *J. appl. Psychol.*, 1953, **37**, 302–307.
3. Wilcoxon, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

The Contribution of Lecture Supplements to the Effectiveness of an Attitudinal Film [1]

Frank T. Staudohar and Robert G. Smith, Jr.

Air Force Personnel and Training Research Center, Lackland Air Force Base

This study had its beginning in the observation of an operational problem in military training. Military training for airmen and officers is highly concerned with changing attitudes and motivations. Hollywood films are frequently available which bear in a general way upon desired changes in attitudes. Yet these films, despite the interest they arouse, are seldom specifically aimed at a given training objective. How can these films be used more effectively to accomplish specific objectives which involve attitude change?

It has been established in many studies that attitudes can be changed by appropriate motion picture films (**2**, **3**, **4**, **5**, **6**, **7**, **8**, **9**, **10**, **11**, **12**). The changes effected by the films frequently have been shown to last over a period of months. At the same time, it has been demonstrated (**4**, **13**) that the effect of the film is likely to be specific, rather than general.

This report describes a study which evaluated one method of making attitudinal films more effective in the accomplishment of specific training objectives. The method involved the use of supplementary lectures in addition to the film. The purpose of the lectures was to stress to the subjects the significant scenes and events in the movie which were related to a specific training objective. Similar lectures have been found to increase the effectiveness of films in teaching facts (**13**).

Method

Subjects. Air Force basic military trainees were used as subjects (*Ss*). They were in the second week of a basic training period of approximately 10 weeks.

Because of limitations of space and military control over the *Ss*, the sampling unit used in the experimental part of the study was the flight, not the individual. A total of 16 basic training flights were used. The number of men per flight averaged 55. There was a total of 876 *Ss* used in the experimental phase of the study, and 311 in a preliminary scale-developing phase.

The film. The commercial motion picture *Twelve O'Clock High* was chosen for this study because it was generally relevant to the development of a favorable attitude toward discipline.

Its story is that of an American Air Force bomber group in England during the early part of World War II. This group was demoralized and suffering heavy losses. Its commander had become too closely identified with his men and could not effectively maintain discipline. He could not bring himself to transfer or punish those who committed serious errors. He is replaced by a general who institutes strict discipline and brings the group up to a highly effective state.

This film was used in the training program of the Officer Military Schools at Lackland Air Force Base.

The lectures. The film was reviewed to determine which scenes were especially relevant to the development of a favorable attitude toward discipline. Lectures designed to point out those scenes particularly relevant to the importance of discipline were developed.

A prelecture pointed out to the *Ss* what to look for in the film. It told of certain scenes to which they should direct their attention and gave a brief resumé of the effects of the variation in discipline as portrayed in the motion picture. A postlecture covered exactly the same topics but informed the *Ss* what they should have noted in the film. Both lectures required about eight minutes.

In addition, two brief lectures were written which included the principal topics of the other two lectures. One of the short lectures was used as an introduction, and the other as a conclusion. These were each about four and a half minutes long and were always used in conjunction with each other.

The lectures were delivered by an Air Force Officer who wore campaign ribbons and pilot's wings. It was felt that the ribbons and wings would increase the prestige of the lecturer as someone who evidently could speak from experience and with authority. The lectures were delivered from notes

[1] This is a summary of a thesis submitted by the senior author in partial fulfillment of the requirements for the degree of Master of Science at Trinity University, San Antonio, Texas. The investigation was carried out under the Air Force Personnel and Training Research Center in support of Project No. 7705. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government.

somewhat informally, although the lecturer confined himself within narrow limits.

The word level of the lectures was evaluated as *Fairly Easy* by means of the Flesch Reading Ease Score. This is estimated as the level which could be read with ease by about 88% of the U. S. population.

The questionnaire. Twenty-three items dealing with discipline were assembled in a tentative attitude scale. This questionnaire was given to 311 Air Force basic trainees. After considering the proportion answering each alternative and the meaning of the item, responses to the items were dichotomized close to the 50% point. Item analyses were performed according to the method of Davis (1). A total of 16 items remained after item analysis (1, Appendix B). An estimate of the internal consistency of the items was obtained by means of Kuder-Richardson Formula 20. The value of the coefficient was .46. Although it is recognized that the Kuder-Richardson procedure underestimates the reliability of the questionnaire, the number of *Ss* was increased to compensate for the low internal consistency.

Procedure. There were four groups with four flights in each group. The assignment of flights to groups was done with the aid of a table of random numbers.

Each of three of the groups heard one of the lectures described above, saw the film, and took the questionnaire. The other group was a control group which saw the film and took the questionnaire. It did not hear any of the lectures.

The questionnaire was administered at the conclusion of the film and/or the lecture. There was no pretest to determine initial attitude toward discipline. It was felt that a pretest might alert *Ss* to those aspects of the film dealing with discipline. If this were true, the effect of the pretest might be sufficiently strong to outweigh the effects of the lectures.

Accordingly, the principle of randomization was resorted to in order to control and equate relevant variables.

Table 1

Significance of Differences in Attitude Toward Military Discipline

Source of Variation	*df*	Mean Square	*F*
Within flights	860	10.15	
Between flights	15	28.53	
Between treatments	3	81.12	7.93*
Between flights within treatments	12	15.47	1.52
Pooled error	872	10.23	

* Significant beyond the .001 level.

Results

The first step in the analysis of the results was analysis of variance. Since the sampling unit was the flight rather than the individual, it was necessary to determine whether significant variation existed between the flights within the treatment groups when compared with the variation between individuals. This comparison did not yield a statistically significant *F* ratio. Accordingly, these two sources of variation were pooled to provide an error term for the comparison of differences between treatments. This comparison was significant at beyond the .001 level. The results of this analysis are shown in Table 1.

The results of the analysis indicated the existence of significant variation in attitude toward discipline, but did not point out which groups differed one from the other. Accordingly, *t* tests were made to compare groups. These results are shown in Table 2.

Every one of the lecture groups resulted in an increase in favorableness of attitude toward discipline as compared to the control group which did not receive a lecture. On the other hand, there were no significant differences between the lectures in their effect on attitude toward discipline.

Discussion

The results of this study clearly indicate that it is possible to improve the attitude-changing character of a motion picture by supplementing it with appropriate lecture presentations.

Table 2

Differences Between Mean Attitude Scores of Experimental and Control Groups

Groups*	Mean Difference	*N*	*SD*	*t*	*P*
I–II	.07	439	.313	.224	
I–III	.03	438	.311	.096	
I–IV	.95	437	.312	3.045	.01
II–III	.04	439	.303	.132	
II–IV	.88	438	.303	2.904	.01
III–IV	.92	437	.302	3.046	.01

* Group designations: I. Pre-post lecture; II. Prelecture; III. Postlecture; IV. Nonlecture.

There are, however, certain limitations that stem from specific practices and materials used in this study. It should not be concluded that a film having only a remote relation to a training objective can be made useful by lecture supplement. *Twelve O'Clock High* contained several sequences bearing very closely on the need for discipline in military forces.

It is also possible that variations in the way the lectures are presented might cause differences in their effectiveness. It may be that, to basic trainees, the officer who delivered the lectures had sufficient prestige that the location of the lectures in relation to the film was not of sufficient strength to overcome his influence.

It should also be mentioned that the *Ss* had only limited experience with the Air Force. It is quite possible that additional experience with military discipline may further change their attitudes toward discipline.

Summary and Conclusions

This report describes an experimental evaluation of the effect on expressed attitudes toward discipline of lecture supplements to a film. The lectures were designed to point out significant sequences in the film which were thought to stress the need for discipline in the military service.

Three lectures were developed. They were comparable in content. One was used prior to the film; another was used after the film; a third was designed to be presented part before and part after the film.

Those *Ss* hearing lectures with the film were generally more favorable in attitude toward military discipline than those who saw the film alone. There were, however, no statistically significant differences between the positions of lectures in their effect on attitude toward discipline.

Received May 18, 1955.

References

1. Davis, F. B. *Item-analysis data.* Cambridge: Harvard Univer. Press, 1949.
2. Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Studies in social psychology in World War II: Experiments on mass communication.* Princeton: Princeton Univer. Press, 1949.
3. Hulett, J. E., Jr. Estimating the net effect of a commercial motion picture upon the trend of local public opinion. *Amer. sociol. Rev.*, 1949, **14**, 263–275.
4. Krech, D., & Crutchfield, R. S. *Theories and problems of social psychology.* New York: McGraw-Hill, 1948.
5. McFarlane, A. M. A study of the influence of the educational geographical film upon the racial attitudes of a group of elementary school children. *British J. educ. Psychol.*, 1945, **15**, 152–153.
6. Murphy, G., Murphy, L. B., & Newcomb, T. M. *Experimental social psychology.* New York: Harper, 1937.
7. Newcomb, T. M. *Social psychology.* New York: Dryden, 1950.
8. Peterson, R. C., & Thurstone, L. L. The effect of a motion picture film on children's attitude toward Germans. *J. educ. Psychol.*, 1932, **23**, 241–246.
9. Peterson, R. C., & Thurstone, L. L. *Motion pictures and the social attitudes of children.* New York: Macmillan, 1933.
10. Raths, L. E., & Trager, F. M. Public opinion and "Crossfire." *J. educ. Sociol.*, 1948, **21**, 345–368.
11. Rosen, I. C. The effect of the motion picture "Gentleman's Agreement" on attitude toward Jews. *J. Psychol.*, 1948, **26**, 525–536.
12. Wiese, M. J., & Cole, S. G. A study of children's attitudes and the influence of a commercial motion picture. *J. Psychol.*, 1946, **21**, 151–171.
13. The Pennsylvania State College. *Practical principles governing the production and utilization of sound motion pictures.* Instructional Film Research Program, 1950.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Leadership and Predictive Abstracting

C. G. Browne and Richard P. Shore

Wayne University

Recently the emphasis of research on leadership by both psychologists and sociologists has been on methods which look upon leadership as a dynamic relationship existing between the "leaders" and the "followers," or between the leader and the group. This is in somewhat sharp contrast with the earlier emphasis in which leaders were studied in isolation from the group, largely in terms of personal characteristics that were measurable within the leader himself (**11**).

In attempting to analyze this dynamic relationship between the leader and the group and to arrive at the variables which may affect the relationship, it has appeared to some that there is at least face validity in the hypothesis that the individual who will be most successful in influencing the actions of a group, and therefore ordinarily classified as the leader, will be the individual who knows the thinking of the group best. With this knowledge the individual can guide his own behavior in attempting to influence the behavior of the group toward some specific action or goal. A number of studies of this "knowledge of others" or "prediction" variable have been published (**1**, **2**, **3**, **4**, **5**, **6**, **8**, **9**, **10**).

Some investigators have referred to the prediction variable as "empathy." To the present authors, however, this is not a satisfactory designation for the phenomenon being studied. The term empathy is borrowed from clinical psychology and carries a connotation which relates to one individual putting himself in another individual's place and in effect experiencing the other person's thoughts and emotions. In short, "to empathize" means something different, perhaps something more, than to predict or to attempt to determine, and just determine, another person's thinking. In addition, empathizing usually is the result of one individual being told by another individual what his thoughts and experiences have been, or at least learning of them through some particular, identifiable channels.

The essence of the leadership variable previously mentioned does not include either (*a*) the taking on of the experience and the emotion of another individual, or (*b*) the necessity of learning specific knowledge regarding another person's experiences through any specific channels. Rather the leadership variable relates to an individual's ability to take from all of his knowledge of individuals or groups those particulars which will prove to be determiners of the behavior or attitude, etc. of an individual or a group on any given question at any given time. Fundamentally the process involved is the process of abstracting, which is the process of selecting certain details from an event and eliminating other details which are included in the same event (**7**). It follows, then, that the extent to which any individual is able to predict the attitudes, etc., of any other individual will be dependent largely on the extent to which he is able to select details from an event which are pertinent to and have an influence on the areas to be predicted. The leadership variable being investigated in this study, therefore, will be referred to as "predictive abstracting" since it is believed that this term is a more satisfactory operational description of the process involved than are the other terminologies which have been used by other investigators.

Specifically, the present study is concerned with the following hypotheses: (*a*) predictive abstracting is a function of leadership; (*b*) a direct relationship exists between an individual's predictive abstracting ability and his echelon level in an industrial organization.

Subjects and Procedure

Four echelon levels of a Detroit metal tubing manufacturing company (Wolverine Tube, Division of Calumet and Hecla, Inc.) were sampled to provide a total of 83 subjects (*Ss*), divided as follows: 5 department managers (most focal echelon used),

9 general foremen selected from various plant departments, 17 assistant foremen from various departments, and 52 nonsupervisory personnel selected at random from an employee roster. The nonsupervisory personnel were sampled proportionately from mill and engineering departments and from each of the three working shifts.

A group-administered questionnaire was devised which consisted of 27 statements, divided into nine statements each for attitudes toward job satisfaction, economic issues, and social issues. The statements were classified into the three areas by agreement of five persons experienced in attitude surveying. Responses were made on a four-point intensity scale ranging from "strongly agree" to "strongly disagree." The relative order of the response columns was reversed at random for approximately half of the statements. On the questionnaire the statements were arranged in random order. Responses were scored as follows: strongly agree, four points; agree, three points; disagree, two points; and strongly disagree, one point. All data were analyzed on the basis of these values.

The questionnaires were administered as follows:

1. Each *S* completed the questionnaire on the basis of his own attitudes toward the questions included. Although the present interest was in the predictive abstracting ability and not in the attitudes as such, it was necessary to obtain a scoring of individual attitudes for comparative purposes with the predictive abstracting responses.

2. Each department manager completed a questionnaire on the basis of his abstracting of the responses of nonsupervisory personnel as a group.

3. Each general foreman and each assistant foreman completed two questionnaires involving predictive abstracting—one for his abstracting of nonsupervisory personnel responses and the other for his abstracting of department manager responses.

4. Each nonsupervisory person completed a questionnaire on the basis of his abstracting of the responses of department managers as a group.

Thus, there were two major types of predictions—those on nonsupervisory personnel made by the three supervisory echelons and those on department managers made by general foremen, assistant foremen, and nonsupervisory personnel.

To determine the accuracy with which one echelon abstracted the attitudes of another echelon, a predictive abstracting score (*PRAB score*) was computed for each statement. The PRAB score is the difference, without regard to algebraic sign, between the value of a predicted response and the actual mean response value of the group on whom the prediction was made. For example, if an individual predicted that the mean response of a group on a given statement would be "strongly agree" or a score of 4, and the actual mean response was 3.1, the PRAB score would be .9. To determine group predictive abstracting accuracy on an item, the mean PRAB scores of all *S*s in the predicting group were used. The *higher* PRAB scores indicate a *less* accurate predictive abstracting ability since a PRAB score of zero would represent perfect prediction.

A nonparametric test of significance, the ranking method (**13**), was used to determine whether two echelons differed significantly in their ability to abstract the responses of another echelon. For example, in comparing the relative ability of department managers and general foremen to abstract the job satisfaction (nine items) of nonsupervisory personnel, ranks ranging from 1 to 18 were assigned to the PRAB scores of each of the predicting echelons. Then the nine ranks for each of the predicting echelons were summed. The smallest sum of ranks was entered in a table of probability values (**12**) and the significance of the difference between the two sets of summed ranks was determined. When a significant difference was obtained, the smaller sum of ranks indicated the echelon which abstracted attitudes with a significantly greater degree of accuracy than the other echelon.

Results

It is hypothesized that predictive abstracting is a function of leadership and that a direct relationship exists between industrial echelon level and PRAB (predictive abstracting) scores. In Table 1 the sums of ranks (based on PRAB scores) of the three predicting echelons on both department managers and nonsupervisory personnel are presented, each echelon being compared with the other two echelons in predictive abstracting ability. In all of the data the smaller sum of ranks indicates the echelon which predicted more accurately.

The data in Table 1 indicate that generally throughout all of the predictive abstracting (*a*) the supervisory personnel predicted more accurately than the nonsupervisory, and (*b*) the more focal supervisory personnel predicted more accurately than the less focal. Although in most cases the differences are not great enough to be statistically significant at the 5% level, the trend is definite enough throughout to justify support of the hypothesis.

In the predictions on the department managers, both the general foremen and the assistant foremen predicted on the total questionnaire with a significantly greater degree of accuracy (1% level) than the nonsupervisory personnel. Although the superiority of predicting was not statistically significant for each of the three divisions of the questionnaire, it was significant at the 1% level for

Table 1

Industrial Echelon Differences in Predictive Abstracting of Attitudes Based on PRAB Scores

Predictive abstracting on department managers	Total Questionnaire	Sum of Ranks: Job Satisfaction	Economic Issues	Social Issues
General foremen	580***	65	64	66
Nonsupervisory	905	106	107	105
Assistant foremen	567***	75	52***	73
Nonsupervisory	918	96	118	98
General foremen	714	94	71	85
Assistant foremen	770	76	100	86
Predictive abstracting on nonsupervisory personnel				
Department managers	659	90	59**	80
General foremen	825	80	112	90
Department managers	671	83	68	83
Assistant foremen	813	88	103	88
General foremen	746	78	95	85
Assistant foremen	739	93	76	85

** Difference between groups significant at 2% level of confidence.
*** Difference between groups significant at 1% level of confidence.
Note.—To be significant, the following sums of ranks are required:

	5%	2%	1%
9 statements	63	59	56
27 statements	629	608	593

the assistant foremen's predictions of economic issues, and approached significance at the 5% level in the predictions of the general foremen for all of the three divisions. The comparisons of the predicting of the general foremen and the assistant foremen yielded no significant differences, either for the total questionnaire or for any of the three divisions.

In the predictions on the nonsupervisory personnel there was a general tendency for department managers to predict the attitudes of nonsupervisory personnel with greater accuracy than either the general foremen or the assistant foremen. Statistical significance was reached, however, only in the predictions of department managers on economic issues when compared with general foremen (2% level). When compared with each other, the general foremen and the assistant foremen were very close in their PRAB scores, and no trend for one group or the other to be superior is indicated.

An hypothesis to the effect that any echelon would predict more effectively the attitudes of individuals in adjoining or near echelons in the organizational hierarchy would not be supported with the present data. In predictions on the department managers, the assistant foremen predicted almost equally well as the general foremen, who are nearer the department managers in echelon level. In predictions on the nonsupervisory personnel, the department managers predicted more accurately than either the general foremen and the assistant foremen, even though both of the foremen groups are nearer the nonsupervisory personnel in the organizational structure and in their daily working contacts. The data would, however, give support to an hypothesis that predictive abstracting is an aspect of an individual's behavior in his relations with other people which is not dependent upon the extent of his acquaintance or contacts with the individuals whose behavior he is predicting.

In Table 2 the data are presented for the comparison of the PRAB scores (in sums of ranks) of the most focal echelon, the depart-

Table 2

Predictive Abstracting of Attitudes by Department Managers and Nonsupervisory Personnel

	Total Questionnaire	Sum of Ranks		
		Job Satisfaction	Economic Issues	Social Issues
Department managers predicting on nonsupervisory	654	99	65	61*
Nonsupervisory predicting on department managers	831	71	106	110

* Difference between groups significant at 5% level of confidence.

ment managers, with the least focal echelon, the nonsupervisory personnel. On the total questionnaire, on economic issues and on social issues, the department managers predicted more accurately, the difference being significant at the 5% level for social issues, and very nearly significant for the economic issues. The nonsupervisory personnel, however, predicted the department managers more accurately on job satisfaction, although the difference is not near significance. These data related to the PRAB scores of the two extreme echelons used in the study again add evidence to the support of the hypothesis that predictive abstracting is related to echelon level, the more focal echelons having higher PRAB scores. From the standpoint of industrial relations, these data may lead to a consideration of the possibility that industrial conflict arises more as the result of lack of understanding of management personnel by nonsupervisory personnel than it does from a lack of management's understanding of nonsupervisory thinking.

The data in Table 3 present comparisons of general foremen and assistant foremen in their predictive abstracting of the attitudes of department managers and nonsupervisory personnel. On the total questionnaire, on economic issues, and on job satisfaction, both foremen groups predicted department managers better than they predicted nonsupervisory personnel, and in all cases the differences were significant at the 5% level or higher. On social issues there is no difference for either foremen group in its predictive abstracting of either the department managers or the nonsupervisory personnel. These data also reject any hypothesis that PRAB scores are likely to be higher when they are related to adjoining or near echelons or if working contacts are closer, since both of the foremen groups are nearer the nonsupervisory personnel in the extent of their contacts with them

Table 3

Predictive Abstracting of Attitudes by General and Assistant Foremen

	Total Questionnaire	Sum of Ranks		
		Job Satisfaction	Economic Issues	Social Issues
Predictive abstracting by general foremen				
Department managers	612*	63*	56**	83
Nonsupervisory	873	108	114	87
Predictive abstracting by assistant foremen				
Department managers	574***	62*	52***	85
Nonsupervisory	911	109	118	86

* Significant at 5% level of confidence.
** Significant at 2% level of confidence.
*** Significant at 1% level of confidence.

and in the organizational hierarchy. In spite of the echelon distance, however, the foremen predicted the attitudes of department managers consistently better. When the data in Table 3 are compared with the data in Table 2 in terms of industrial relations they may support the beliefs of many business executives that the great need for training in human relations and general understanding of human problems is with the outer echelons of supervision, with those individuals who are closest to nonsupervisory personnel but who show lack of understanding in dealing with them.

Summary

Eighty-three employees of Wolverine Tube, a Detroit manufacturing firm, representing four echelons of the business (department managers, general foremen, assistant foremen, and nonsupervisory personnel) were administered an attitude questionnaire dealing with job satisfaction, economic issues, and social issues. Following this, each echelon predicted the attitudes of the department managers as one group and the nonsupervisory personnel as a second group. The process involved in these predictions is designated as "predictive abstracting" and the difference between the individual's predictions and the mean score of the group on whom he is predicting is designated as a PRAB score. On the basis of the predictive abstracting of the various echelons, PRAB scores were obtained for the department managers' predictions on the nonsupervisory personnel, the nonsupervisory personnel on the department managers, and the general foremen and the assistant foremen on both the department managers and the nonsupervisory personnel. A nonparametric test of significance, the ranking method, was used in analyzing the data.

In this study it is assumed that leadership is involved in the performance of supervisory and executive industrial functions. Throughout all of the predictive abstracting, the data generally support the hypothesis that predictive abstracting is an aspect of leadership, with the following specific observations: (*a*) the supervisory personnel predicted more accurately than the nonsupervisory, and (*b*) the more focal supervisory echelons predicted more accurately than the less focal. A study is needed which will test the extent to which predictive abstracting is a function of the individual and/or a function of the position which the individual occupies. When these data are available, we shall be in a better position to determine whether or not the PRAB score of an individual may be used effectively in predicting his probable performance as a leader or supervisor or executive, or whether the PRAB score will serve only as an indicator of areas of training which are needed if the individual is to perform successfully in a leadership capacity.

Received April 4, 1955.

References

1. Bender, I. E., & Hastorf, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556–561.
2. Browne, C. G. Study of executive leadership in business. III. Goal and achievement index. *J. appl. Psychol.*, 1950, **34**, 82–87.
3. Browne, C. G., & Neitzel, Betty J. Communication, supervision, and morale. *J. appl. Psychol.*, 1952, **36**, 86–91.
4. Chowdry, K., & Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
5. Dymond, Rosalind F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.
6. Kerr, W. *The empathy test.* Chicago: Psychometric Affiliates, 1947.
7. Korzybski, A. *Science and sanity.* Lakeville, Conn.: Institute of General Semantics, 1948.
8. Norman, R. D., & Ainsworth, Patricia. The relationships among projection, empathy, reality, and adjustment, operationally defined. *J. consult. Psychol.*, 1954, **18**, 53–58.
9. Remmers, L. J., & Remmers, H. H. Studies in industrial empathy. I. Labor leaders' attitudes toward industrial supervision and their estimate of managements' attitudes. *Personnel Psychol.*, 1949, **2**, 427–436.
10. Remmers, L. J., & Remmers, H. H. Studies in industrial empathy. II. Managements' attitudes toward industrial supervision and their estimate of labor attitudes. *Personnel Psychol.*, 1950, **3**, 33–40.
11. Stogdill, R. M. Personal factors associated with leadership: a survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.
12. Wilcoxon, F. Probability tables for individual comparison by ranking methods. *Biometrics*, 1947, **3**, 119–122.
13. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford: American Cyanimid Company, 1949.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Validity of Extrapolating Nonresponse Bias from Mail Questionnaire Follow-ups [1]

Herbert Zimmer [2]

Air Force Personnel and Training Research Center, Officer Education Research Laboratory, Maxwell Air Force Base, Alabama

Sample surveys usually seek to describe some aspect of a defined population. As such, their accuracy rests on the precision with which the sample represents the population. A considerable body of literature points to the possibility that nonrespondents may differ markedly from respondents in any survey sample, and that neither group may constitute an unbiased sample of the population by itself. The literature on nonresponse bias in mail questionnaires is far too extensive to be discussed here, but many of the relevant articles are summarized in two reviews (**7**, **9**). The problem of nonresponse bias is not restricted to mail questionnaires alone, but occurs in any survey that attempts to represent a population fairly, regardless of the research tool employed (**4**). An awareness of this possible bias has led surveyors to use a number of ingenious devices for increasing the initial response and for follow-up communications, in order to reduce the proportion of nonrespondents. Successful as such efforts usually are, the careful investigator is still obliged to describe the inevitable residue of nonrespondents, inasmuch as they continue to pose the possibility of prejudicing the sample.

The surest and most adequate method of accounting for nonrespondents to mail questionnaires is by personal contact and interview. In practice this is rarely done, since even the pursuit of only a sample of nonrespondents by telephone or intensive mail barrage is costly and time consuming (**1**, **4**). Several investigators (**2**, **3**, **8**) have attempted to devise a shortcut to obtaining the unknown characteristics of nonrespondents. They employed, or suggested, the use of the several coordinate values of the initial response group and of follow-up response groups to extrapolate any obtained trend to unknown values for the group of nonrespondents.

A rationale for this procedure can be elaborated. The successive subgroups within a sample which range from the initial response group, through the various groups which responded to an ever-increasing number of follow-up communications, to the group of nonrespondents may be said to have a decreasing probability of response and an increasing probability of nonresponse. This relationship may be expressed symbolically, beginning with the initial response group and ending with the nonresponse group, as

$$p^n q^n + p^{n-1} q^{n+1} + p^{n-2} q^{n+2} + p^{n-3} q^{n+3} \ldots\ldots + p^{n-(n-1)} q^{n+(n-1)} + q^{n+n}$$

For the residual nonresponse group (q^{n+n}), the probability of nonresponse is given as unity. The above expression is suggestive of one-half of the binomial expansion. The fitting of curves derived from empirical data can determine the applicability of this formulation.

Any factor which varies systematically with the probability of response-nonresponse may presumably be extrapolated to the nonresponse group from the set of coordinates furnished by the successive groups of respondents. Even on purely theoretical grounds this does not imply that all factors can be accurately predicted. Some factors may be unrelated to the response-nonresponse variable, others may be subject to large random variations; for still others the trend curve may contain unexpected inflections in the area to be extrapolated. If a factor is systematically related to the response-nonresponse variable, curves which represent this relationship and which have been derived from a number of independent samples would be expected to

[1] The author is indebted to Harry M. Henkin and Stephen J. Zolczynski for their help in carrying out the statistical computations.

[2] Now at Georgetown University School of Medicine.

follow a common pattern. By the same token, the absence of a common pattern for any particular factor would indicate the lack of a stable relationship, which would eliminate the response-nonresponse continuum as an adequate basis for projecting that factor to the nonresponse group.

A definitive answer to the question of which and how many variables can be extrapolated with what degree of certainty to a nonresponse group must come from an evaluation of the identical set of factors for a series of samples. This is a research task of considerable proportions. The present experiment is of a more limited scope. Nevertheless, it can provide a test of the fundamental hypothesis.

Hypothesis. The response-nonresponse probability function indicates the presence and direction of nonresponse bias.

Procedure

During the first two weeks of February 1954, a questionnaire was sent by registered mail, with return receipt requested, together with an explanatory letter and return envelope, to 220 U. S. Air Force officers and airmen. The questionnaire concerned an important life experience in which each of the 220 addressees had recently participated. About one month after the mailing of the questionnaires, 166 of the questionnaires had been returned. At that time, early in March, a follow-up letter was sent to the 54 nonrespondents. Between the mailing of the follow-up letter and the first week in April an additional 26 questionnaires were returned. Four of these questionnaires were returned so soon after the

Table 1

Distribution of High-, Medium-, and Low-Response Probability Groups on Seven Variables

Characteristic		Group: High ($N=170$)	Medium ($N=22$)	Low ($N=28$)	All Subjects ($N=220$)
1. Age	Mean	29.00	27.84	26.82	28.61
	SD	4.06	4.06	5.43	4.33
2. Education	Mean	14.73	13.91	13.71	14.52
	SD	1.69	2.06	2.09	1.83
3. Rank	Mean	8.20	7.77	6.41	7.93
	SD	2.10	2.19	3.07	2.34
4. Years of service	Mean	6.55	5.33	5.35	6.28
	SD	3.27	3.28	3.53	3.35
5. Marital status					
Single	*f*	54	6	15	75
Married	*f*	100	15	10	125
Undetermined*	*f*	16	1	3	20
6. Military status					
Regular	*f*	60	7	15	82
Reserve	*f*	109	15	11	135
Undetermined*	*f*	1	0	2	3
7. Occupation					
Pilot and copilot	*f*	91	12	8	111
Bombardier and navigator	*f*	22	1	4	27
Gunner	*f*	29	5	8	42
Radio, radar and recon specialists	*f*	14	2	2	18
Crew chief	*f*	8	1	2	11
Other	*f*	6	1	4	11

* Not included in computations for Table 2.

Table 2

Group Differences Between High-, Medium-, and Low-Response Probability Groups on Seven Variables

	Differences between Groups											
	Over-all			High vs. low			High vs. medium			Medium vs. low		
Characteristic	*df*	*F*	*p*	CR		*p*	CR		*p*	*t*		*p*
1. Age	2/217	3.50	<.05	2.00		<.05	1.62		<.15	.75		<.50
2. Education	2/217	5.21	<.01	2.41		<.02	1.75		<.10	.33		>.70
3. Rank	2/217	7.40	<.001	2.92		<.01	.86		<.40	1.79		<.10
4. Years of service	2/169	1.83	—	—		—	—		—	—		—
	df	χ^2	*p*	*df*	χ^2	*p*	*df*	χ^2	*p*	*df*	χ^2	*p*
5. Marital status	2	6.50	<.05	1	5.64	<.02	1	.35	>.50	1	4.54	<.05
6. Military status	2	5.08	<.10	1	4.68	<.05	1	.11	>.70	1	3.21	<.10
7. Occupation	10	12.03	>.20		—			—			—	

mailing of the follow-up letter that the respondents could not have had time to receive the follow-up letter before returning their questionnaires. These four questionnaires were therefore grouped with the 166 earlier respondents, to give a total of 170 respondents, and 22 late respondents. Eleven months after the mailing of the questionnaires, 28 addressees still had not responded. Of these 28 questionnaires, two were returned by the Post Office as being undeliverable or unacceptable to the addressee. All envelopes carried a legible return address and were registered mail.

This results in the following breakdown by number of subjects in three categories.

1. Questionnaires returned before receipt of follow-up letter by respondent: high response probability 170
2. Questionnaires returned after receipt of follow-up letter by respondent: medium response probability 22
3. Questionnaires not returned: low response probability 28

Total questionnaires mailed 220

A number of personal history items were available in military records for most of our subjects. The three groups, high response probability group ($N = 170$), medium response probability group ($N = 22$), and low response probability group ($N = 28$), were compared with each other on seven variables: (*a*) age at last birthday, (*b*) education, (*c*) military rank, (*d*) years of military service, (*e*) marital status, (*f*) military status, and (*g*) military occupational specialty.

Results

The first four of these variables can be represented as numerical distributions. Their means and standard deviations are shown in Table 1. Frequencies are shown for the last three variables, since information for these variables is of an unordered nature.

On the first four variables the medium response probability group deviates from the high response probability group in the same direction as the low response probability group. On three of these four variables the medium response probability group occupies an intermediate position between the two other groups. On the remaining variable, years of service, analysis of variance does not indicate a significant difference between the three groups, as shown in Table 2.

Table 2 shows the statistical significance of differences between the three groups. For each variable, an over-all test was done first. If it suggested that the three groups were not likely to be random samples from a common population, then the three groups were compared with each other, two at a time. At the top of Table 2 appear the four variables which yielded numerical distributions. For these, analysis of variance was used for the over-all test, CR and *t* for the group-by-group comparisons. The lower part of Table 2 shows the three variables for which categorical information was available. Chi-square tests were applied to these data.

Variable 7, military occupational specialty, shows no significant differences between groups, even though the small cell frequencies lead to an overestimation of existing differences.

The first three variables shown in Table 2

—age, education, and rank—are consistent with the hypothesis, though owing, perhaps, to the limited number of cases, all of the group-by-group comparisons do not attain significance.

Variables 4 and 7, years of service and military occupational specialty, are also consistent with the hypothesis. For these two variables there is no significant nonresponse bias, and none is indicated by the response-nonresponse probability function.

Variables 5 and 6 are not consistent with the hypothesis. While a significant nonresponse bias exists here, the medium response probability group fails to indicate this, and is itself closer to the high response probability group than to the low response probability group.

Of the seven variables against which the hypothesis was evaluated, five are found to be consistent with it, and two not consistent with it. Inasmuch as these tests of the hypothesis are based on the same sample and may not be entirely independent of each other, it is not feasible to name the odds of getting this particular combination of agreements and disagreements with the hypothesis.

Discussion

The present data give some support to the hypothesis that the response-nonresponse probability function indicates the presence and direction of nonresponse bias. Any acceptance of this hypothesis must be qualified by the obvious exceptions to it. Though this method offers the advantage of being more economical and less time-consuming than alternative methods, it should be recognized that it is also likely to be less accurate, which may more than offset its advantages. Indications from this study and from similar studies (**2**, **6**) suggest that the use of the method considered here may introduce errors in estimating the extent and even the presence of nonresponse bias for some variables. As stated above, larger studies are needed to establish more conclusively the predictability of nonresponse bias by means of this method. In the meantime, exclusive reliance on this method to account for all possible aspects of nonresponse bias must be suspect.

Evidence from the few other studies available on this topic show remarkable agreement with the present results. Edgerton, Britt, and Norman (**5**) in a mail questionnaire study of contestants in the First Annual Science Talent Search report data showing a step-wise decrease in mean Science Aptitude Examination scores for three groups, totaling 906 subjects: initial respondents, respondents to a follow-up letter, and nonrespondents. Baur (**2**) reports the results of a mail survey by the Veterans Administration of 6,000 veterans eligible for education under the G.I. Bill. Data on five variables are given for five separate groups: early initial respondents, late initial respondents, respondents to first follow-up letter, respondents to second follow-up letter, and nonrespondents. The groups show a gradual and progressive decrease in formal education and in the proportion of subjects with definite educational plans, though data for nonrespondents are not available on the educational plans variable. No significant difference between groups is reported for age and for parenthood. The groups show a gradual increase in the proportion of married subjects, with a reversal of this trend by nonrespondents, who are closest to the early initial respondents on this variable. Ford and Zeisel (**6**) analyzing returns from a mail questionnaire to 382 former employees of a company, in terms of initial respondents, respondents to first follow-up, and respondents to second follow-up, found that nonrespondents actually had a somewhat lower proportion of employees rated unsatisfactory by their supervisors than these investigators were led to expect by the highly consistent trend established from the three other groups. Phillips (**10**) compared initial respondents, follow-up respondents, and nonrespondents in a group of 93 Fisk University alumni on four factors—year of graduation, sex, marital status, and parenthood—and reports no statistically significant differences between his groups on any of these four factors. As does the present experiment, these studies demonstrate the predictive ability of the response-nonresponse probability function for a number of variables, and also its failure in the case of other variables.

Owing to availability of data on nonrespondents, experimental evidence is mostly limited to a number of characteristics which as a rule are not themselves the topic under study in mail questionnaires, but there is every indication that questionnaire items show a similar distribution along the response-nonresponse continuum (6).

Summary

This experiment attempted to study the accuracy with which nonresponse bias in mail questionnaires can be extrapolated from the trend derived from the several coordinate values of initial and follow-up response groups for any given variable. This method is much more economical of time and effort than an interview sampling of nonrespondents. Data on seven variables were obtained for three groups, totaling 220 subjects and representing three points on a response-nonresponse probability dimension. These data were employed to evaluate the hypothesis that the response-nonresponse probability function indicates the presence and direction of nonresponse bias. The results indicate that of the seven variables considered, five were found to be consistent with the hypothesis and two not consistent with it. Furthermore, nonresponse bias occurred despite a very high rate of return. These results were in keeping with the findings of other investigators. Large-scale, definitive studies on the predictability of nonresponse bias for all relevant variables by means of this method have yet to be conducted. It was concluded that in the absence of such studies, exclusive reliance on this method to account for all possible aspects of nonresponse bias must be suspect.

Received June 20, 1955.

References

1. Barnette, W. L. The non-respondent problem in questionnaire research. *J. appl. Psychol.*, 1950, **34**, 397–398.
2. Baur, E. J. Response bias in a mail survey. *Pub. Opin. Quart.*, 1947, **11**, 594–600.
3. Clausen, J. A., & Ford, R. N. Controlling bias in mail questionnaires. *J. Amer. statist. Ass.*, 1947, **42**, 497–511.
4. Deming, W. E. On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse. *J. Amer. statis. Ass.*, 1953, **48**, 743–772.
5. Edgerton, H. A., Britt, S. H., & Norman, R. D. Objective differences among various types of respondents to a mailed questionnaire. *Amer. sociol. Rev.*, 1947, **12**, 435–444.
6. Ford, R. N., & Zeisel, H. Bias in mail surveys cannot be controlled by one mailing. *Pub. Opin. Quart.*, 1949, **13**, 495–501.
7. Norman, R. D. A review of some problems related to the mail questionnaire technique. *Educ. psychol. Measmt*, 1948, **8**, 235–247.
8. Pace, C. R. Factors influencing questionnaire returns from former university students. *J. appl. Psychol.*, 1939, **23**, 388–397.
9. Parten, Mildred. *Surveys, polls, and samples.* New York: Harper, 1950.
10. Phillips, W. M. Weaknesses of the mail questionnaire. *Sociol. soc. Res.*, 1951, **35**, 260–267.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

A Biographical Inventory for Students: II. Validation of the Instrument

Laurence Siegel

Miami University [1]

The first article in this series described the development and standardization of a self-administering biographical inventory suitable for administration to male high school seniors and college freshmen (**6**). This Biographical Inventory for Students (BIS) yields scores on 10 subscales which are relatively homogeneous, independent, and reliable. These subscales are designated: *Act*—Action; *Soc*—Social Activities; *Het*—Heterosexual Activities; *Rlg*—Religious Activities; *LMA*—Literature, Music, and Art; *Pol*—Political Activities; *SEL*—Socioeconomic Status; *Eco*—Economic Independence; *Dep*—Dependence upon the Home; *Con*—Social Conformity.

Problem

This paper describes a series of investigations designed to ascertain the usefulness of the BIS as an adjunct to the counseling of students comparable to the standardization groups. Since the BIS items are factual in nature, they appear to be distinctly superior to items of the "I think . . ." or "I feel . . ." variety, thereby combining some of the advantages of personal history forms with those of personality inventories. Validation of the 10 BIS subscales, however, is prerequisite to appraisal of the usefulness of this instrument.

The investigations to be described related the subscales to four classes of criteria: scholastic ability and achievement, personality inventories, values inventories, and vocational choice. Most of these criteria were derived from performance on psychometric instruments. Although such an approach to the validation of the BIS lacks finality, it does permit for some understanding of the interpretation to be accorded the subscale scores.

Procedure

Data for the entire series of studies were obtained from the responses of three basic samples of college freshmen.

Sample I: Entering freshmen at the State College of Washington ($N = 154$). The BIS was included in a battery of instruments administered routinely during freshman orientation week.

Sample II: Freshmen who had been in residence at the State College of Washington for one month prior to administration of the BIS ($N = 334$). These students were enrolled in a physical education course and were required to complete a variety of psychometric instruments as part of an integrated research program.

Sample III: *Superior* freshmen at The Ohio State University ($N = 66$). The bases for classification of these students as "superior" consisted of placement at the seventy-fifth percentile or higher on The Ohio State Psychological Examination, high school grades in the upper third of their graduating class, and a history of participation in extracurricular activities while attending high school.[3]

The research design is schematized in Table 1, which summarizes the criterion variables and the sample or subsample employed for each analysis.

Results

The statistically significant findings resulting from these investigations (with the exception of those utilizing vocational choice as the criterion) are summarized in Table 2. Any interpretation of these findings must be tempered by considerations of size and representativeness of the sample.

Criteria of Ability and Achievement

Although the BIS was not intended to be a test of scholastic ability, it seemed reasonable to expect that some of the subscales might correlate with achievement as defined by course grades. As expected, the BIS subscales (with the exception of *Act*) are not useful for the prediction of ability as defined by ACE total score. Three of the subscales,

[1] A large portion of the research described in this paper was completed while the writer was on the faculty of the State College of Washington, and on the staff of the Personnel Research Board of The Ohio State University.

[3] Analyses based upon this sample were conducted by the staff of the Occupational Opportunities Service, The Ohio State University.

however, are significantly correlated with the criterion of GPA.

The regression of *Dep* upon GPA was curvilinear (eta = .48) with a tendency for high grades to be earned at both ends of the *Dep* continuum. If we accept the premise that students earning mid-range *Dep* scores are in the process of establishing their independence from familial influence, it then follows that such students would neither be as secure as those who have already completed the process of "psychological weaning" nor as secure as those who have not yet started on it. Such insecurity might be expected to result in low grades. Of course, such an interpretation is extremely tentative and awaits further substantiation.

Another study relating BIS subscales to measures of intellectual ability and achievement was conducted with Sample III. Although these data are based upon a highly selected sample constituting one tail of the distributions of ability and achievement, there are two points of agreement between this study and the one discussed previously: *Act* correlates most significantly (in a negative direction) with the test of ability; and *Het* correlates negatively with the criteria of grades. This latter correlation is of particular interest in view of the finding that the correlation between OSPE and PHR for this truncated sample was only .33.

The relationships cited in Table 2 are congruent with logical expectations. The better students appear to participate in fewer heterosexual, physical, and social activities than do the less capable. In summary, the picture is one of scholastic achievement obtained, to some extent, at the expense of nonintellectual pursuits.

Personality Inventories

Two personality inventories served as validation criteria for the BIS. The first of these, the Heston Personality Inventory (3) provides six scores designated:

Analytical Thinking—"thinking introversion" and "intellectual independence";

Sociability—making friends easily and taking the lead in social participation;

Emotional Stability—ability to remain in stable and uniform spirits and to relax and avoid tension;

Table 1

Summary of Criterion Variables and Samples

Criterion	Sample (or Subsample)	*N*
1. *Scholastic ability and achievement*		
American Council on Education Examination (ACE)	I	154
Grade-point average (GPA) after two semesters	I	154
Ohio State Psychological Examination (OSPE)	III	66
Point-hour ratio (PHR) at end of first quarter	III	66
Rank in high school graduating class	III	66
2. *Personality Inventories*		
Heston Personal Adjustment Inventory	IIA	107
Runner Personality Analysis	III	66
3. *Values Inventories*		
Allport-Vernon Study of Values	IIB	85
Lurie Values Inventory	III	66
4. *Vocational Choice*		
Major field of study	IIC	185

Confidence—ability to make decisions readily and confidently;

Personal Relations—feeling that other people are trustworthy and congenial;

Home Satisfaction—pleasant family relations.

All but one of the obtained correlations were compatible with our prior expectations. The unexpected relationship was that between LMA and Sociability. Perhaps the most parsimonious interpretation in the absence of corroborating evidence is to consider this correlation a result of the type of sampling error liable to occur whenever all variables in a matrix are intercorrelated. In addition, two relationships were hypothesized but were not substantiated by the data: *Het* with Confidence, and *Soc* with Personal Relations.

The curvilinear regression of *Dep* upon Heston's Confidence score yielded an eta of .55. Low *Dep* scores tended to be associated with mid-range Confidence scores, whereas high *Dep* scores were associated with the two extremes of the Confidence scale. The find-

Table 2

Summary of Significant Correlations Between the BIS and External Criteria

Criterion	*Act*	*Soc*	*Hel*	*Rlg*	*LMA*	*Pol*	*SEL*	*Eco*	*Dep*	*Con*
Ability and Achievement										
ACE	−.22***									
GPA (after 2 sem.)	−.25***		−.25***			.25***			(.48)†	
OSPE	−.31**	−.30**	−.26*						−.28*	
PHR (after 1 qtr.)		−.36***	−.45***							
H.S. rank			−.34***					−.34***		.28*
Personality Inventories										
Heston:										
Analytic					.38***	.30***				
Sociability		.56***	.38***		.28***	.36***	.21*			
Confidence	.23*	.20*						.25**	(.55)†	
Home Satisfaction									.22*	
Runner:										
Conventionalism				.34***					.32***	.31**
Practical Attitude							.32**			
Methodical						.28*				.32**
Thing Interest								.34***		
Human Welfare				.29*						
Sociability	.44***	.54***	.53***						.30*	
Exhibitionism		.33**	.30*							
Affectional Expression		.31**							.25*	
Dependence		.32**								
Wish for Dominance	.36***	.60***								
Aggressiveness	.28*	.33**	.34***				.29*		.24*	
Hostility							.27*			
Values Inventories										
Allport-Vernon:										
Aesthetic					.25*					
Social										
Political	.27*	.26*				.31**				
Religious				.38***						
Economic										
Theoretical				−.32**						
Lurie:										
Aesthetic					.45***					
Social	.26*			.32**		.26*			.31**	.27*
Political		.44***				.30*				
Religious	.28*			.48***					.26*	.31**
Economic		.31**								
Theoretical										

* $p < .05$.
** $p < .02$.
*** $p < .01$.
† eta.

ing that students who are overly dependent upon the home (high *Dep* score) may either lack confidence or be overly confident seems quite plausible. If we assume that verbalizations of lack of confidence or of overconfidence may be symptomatic of poor adjustment, we might then expect overdependence to be part of the total configuration. Such an

expectation does not imply causality. Indeed, the lack of confidence (either verbalized or camouflaged by bravado) and the dependent relationship may both be reflections of a more basic causal agent.

The second criterion of personal adjustment was the Runner Personality Analysis (5) which provided 13 criterion scores defined as:

Conventionalism—response to what *ought* to be;
Practical Attitude—concern with utilitarian and material values;
Methodical—pre-planned, systematic performance of activities;
Things Interest—pleasure in hand skill activities;
Human Welfare—the desire to be of social service;
Sociability—desire to be with people;
Exhibitionism—desire to be the center of attention;
Affectional Expression—the need for warm personal relationships with others;
Dependence—the need for help and attention from others;
Wish for Dominance—love of power and authority over people;
Competitiveness—stimulation of effort by the desire to win out over others;
Aggressiveness—the active expression of dissatisfaction with other people;
Hostility—passive resentment; distrust of others.

This analysis was conducted without the formulation of prior hypotheses in view of the fact that the basis of interpretation of the Runner is the configuration of subtest scores rather than any given score per se. Nevertheless, a number of significant correlations which are readily explicable, and might have been predicted, appear in the matrix.

Values Inventories

The BIS subscales were related to scores on the Allport-Vernon Study of Values (1) and the Lurie Values Inventory (4). Both of these criterion instruments yield scores on the six variables classified by Spranger (7). Four consistent relationships appear in both the Allport-Vernon and the Lurie: *Soc*—Political Values; *Rlg*—Religious Values; *LMA*—Aesthetic Values; and *Pol*—Political Values. Aside from these consistencies, it appears that the BIS bears greater resemblance to the Lurie than to the Allport-Vernon.

Vocational Choice

The 334 students in Sample II were classified on the basis of major field of study. Three of the subgroups-by-major contained more than 30 subjects and were therefore considered suitable for further analysis: Business Administration ($N = 34$), Agriculture ($N = 51$), and Engineering ($N = 97$). Mean BIS subscale scores for these groups were compared by computation of t ratios for significance of difference. Three of the subscales differentiated between these groups. (*a*) Business Administration majors, as a group, scored higher than Agriculture majors ($p < .001$) and Engineering majors ($p < .05$) on *Soc*. Furthermore, Engineering majors scored significantly higher on this subscale than did majors in Agriculture ($p < .01$). (*b*) Students majoring in Agriculture scored higher on *Rlg* than did Engineering students ($p < .001$). (*c*) Majors in Business Administration had a higher mean score on *LMA* than majors in Agriculture ($p < .05$). The direction of these differences lends support to the validity of the instrument.

Discussion

We have cited correlations between the BIS subscales and 10 criteria, four of which yielded multiple scores. Collation of the validities obtained for each subscale will serve to clarify the significance and interpretation of scores on the BIS.

Act—Action. This subscale is negatively related to intellectual ability and achievement, except in the case where PHR for a truncated sample served as the criterion. High *Act* scores correlate with several criteria which may be broadly classified as "social" and "striving": i.e., Confidence on the Heston; Sociability, Wish for Dominance, and Aggressiveness on the Runner; Political on the Allport-Vernon.

The significant correlation between *Act* and Religious Values on the Lurie is inexplicable in the light of the total configuration.

Soc—Social Activities. Although this scale does not correlate with either intellectual ability or achievement when the full criterion range is available, it is negatively related to

these criteria for the sample of superior students. It appears that the best of the superior students either are not interested in or haven't the time for such social activities.

Soc scores are strongly correlated with the indices of interest in people, either as ends in themselves or as means to gratification of power strivings, provided by all of the criterion instruments. The validity of this subscale is further supported by the fact that the students majoring in Business Administration score extremely high when compared to students majoring in agriculture and engineering.

Het—Heterosexual Activities. Scores on this variable correlate negatively with criteria of scholastic achievement and positively with criteria of "sociability" as defined by the Heston and Runner. Furthermore, this subscale is related to Runner's Exhibitionism and Aggressiveness scales.

Rlg—Religious Activities. Significant correlations appeared when *Rlg* was related to indices of attitudes which might be classified as "moralistic" and "interest in social welfare activities"; e.g., Conventialism and Human Welfare on the Runner, Religious values on both the Allport-Vernon and the Lurie, and Social values on the Lurie.

The negative relationship between *Rlg* and Theoretical values on the Allport-Vernon is somewhat substantiated by the fact that Engineering majors scored lower on this subscale than did majors in Agriculture.

LMA—Literature, Music and Art. The primary criteria for validation of this subscale was "Aesthetic Values" as defined by the Allport-Vernon and the Lurie. Significant correlations were obtained in both cases. These validities are supported by the fact that *LMA* correlated with Heston's Analytic scale (which is defined as "thinking introversion").

Pol—Political Activities. This BIS subscale correlates with Political Values as defined by both the Allport-Vernon and the Lurie. It is also related to assorted criteria of a methodological and analytic approach to problems.

SEL—Socioeconomic Status. Contrary to our prior hypotheses, *SEL* did not correlate with Economic values. However, this subscale is clearly related to concerns of utility and practicality as well as to caution and adherence to tradition as measured by the Runner.

A convincing demonstration of the validity of *SEL* scores was cited in the first paper in this series (**6**) wherein college freshmen were reported to score significantly higher than high school seniors.

Eco—Economic Independence. Our original expectation was that this subscale would correlate with Economic values of the type measured by values inventories. Since this expectation was not supported by the data, we must conclude that economic independence as measured by the BIS is unrelated to placing high value upon monetary goals. However, this subscale does yield readily explicable correlations with Heston's Confidence score and Runner's index of pleasure in hand-skill activities.

Dep—Dependence upon the Home. Several relationships between this predictor and criteria of so-called "wholesome adjustment" were obtained: i.e., Home Satisfaction on the Heston; Conventialism, Affectional Expression, and Sociability on the Runner; Religious and Social values on the Lurie.

Con—Social Conformity. Scores on *Con* are related to indices of a conventional and methodical approach to matters as defined by the Runner, and to altruistic values (Social and Religious) as measured by the Lurie.

Received October 1, 1954.

References

1. Allport, G. W., & Vernon, P. E. *A study of values.* Boston: Houghton Mifflin, 1931.
2. American Council on Education. *Psychological examination for college freshmen.* Princeton: Educational Testing Service, 1948.
3. Heston, J. C. *Personal adjustment inventory.* Yonkers, N. Y.: World Book, 1949.
4. Lurie, W. A. A study of Spranger's value types by the method of factor analysis. *J. soc. Psychol.*, 1937, **8**, 17–37.
5. Runner, Jessie R., & Runner, K. A gestalt analysis of personality. Unpublished manuscript, Iowa City, 1952.
6. Siegel, L. *A biographical inventory for students: I. Construction and standardization of the instrument. J. appl. Psychol.*, 1956, **40**, 5–10.
7. Spranger, E. *Types of men.* New York: Stechert, 1928.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Equivalence of Forms of the Wonderlic Personnel Test: A Study of Reliability and Interchangeability

H. B. Weaver
University of Hawaii
and C. A. Boneau
Duke University

The Wonderlic Personnel Test is one of the most widely used group tests of general intelligence. It was designed especially for industrial use, is short (12-minute time limit), virtually self-administering, and easy to score. Industrial norms are available on approximately 37,000 cases. It is published in five forms, A, B, D, E, and F—Forms D, E, and F consisting of items selected from the Otis Higher, Forms A and B consisting of similar items but of a wider variety.

The five forms are claimed by the author of the test to be equivalent and interchangeable, and the published norms accordingly are not differentiated by form. This paper presents evidence, however, that there are substantial differences in difficulty among the forms and that they are by no means interchangeable.

I

The study began with a general impression that Form A was yielding higher scores than Form B in routine use in a large industrial concern. A check on the records of nearly 300 cases in which the two forms had been given at random to supervisory candidates showed that Form A seemed to be distinctly easier than Form B, the mean difference being 3.3 score points. This corresponds to a difference of from 10 to 20 centile units in the average range, varying somewhat with the norms used.

II

This suggestive finding led to the following study: 30 students comprising a class in psychology were given both Form A and Form B, half getting A first and half B first. It was found that the group averaged 1.97 more correct answers on Form A than on B. The difference yields a t ratio of 2.61, which is significant at the 2% level for 29 df. The previous results were therefore confirmed.

III

Although these independent findings rather clearly indicated that Form A is easier than B, the question of differences in difficulty among all five forms remained unanswered. Accordingly a further study was undertaken. Seventy *Ss*, comprising two classes in psychology, were given all five forms in counterbalanced order as shown in Table 1. This "unit" design for five *Ss* was replicated 14 times, making a total of 70 *Ss*.

Analysis of the resulting data by forms yielded the mean scores and *SD*'s shown in Table 2. It will be noted that the difference between Forms A and B is 1.95, almost exactly the same as the 1.97 obtained previously on a similar sample. The maximum difference is 5.07, between Forms B and F. The average difference is 2.34.

Inspection of the means and *SD*'s of Table 2 seems to indicate that the forms fall roughly into two groups, A and B comprising a group of greater difficulty and higher variability than D, E, and F. This accords with the history of the development of the test. Forms D, E, and F are made up of items selected from the Otis Higher, while A and

Table 1

Paradigm for Equalizing Practice and Fatigue Effects Among the Five Forms A, B, D, E, F

Order of Administration	Subject				
	1	2	3	4	5
First	A	B	D	E	F
Second	B	D	E	F	A
Third	D	E	F	A	B
Fourth	E	F	A	B	D
Fifth	F	A	B	D	E

Table 2

Means and *SD*'s of Scores on the Five Forms for 70 *Ss*

Form	A	B	D	E	F
Mean	29.79	27.84	31.37	31.31	32.91
SD	6.01	7.39	5.73	5.57	5.7¢

B were developed later and include types of items not found in the Otis.

The data were subjected to analysis of variance as shown in Table 3. Individual differences were, of course, highly significant, yielding an F of 10.92. Form differences yielded an F of 20.15, also highly significant.

Each of the 10 possible differences between forms was tested individually by computing a t ratio based on the error variance estimate of Table 3. The differences and the results of the t tests are shown in Table 4. Nine of the 10 differences are significant, two at the 2% level, three at the 1% level, and four at the .1% level.

It may be noted that the differences are not merely statistically significant but are also of practical significance because of the nature of the test and the recommended manner of its use. The Wonderlic is very commonly used as a screening device, and critical scores are set for employment and other purposes. If the forms are used interchangeably it would therefore appear that meeting or failing to meet a given minimum score would in many cases depend on the particular form an applicant chanced to be given. This chance factor is as high as 5 points, which is 10% of the total possible score range, and in centile units of the published norms may amount to 37% of the distribution (**4**, Table 1, Norms for male college students). The average interform difference, 2.34, is 4.7% of the total possible score range and in centile units may be as much as 20% of the distribution. In view of these variations in difficulty, separate tables of norms should be constructed for each form, or conversion formulas should be developed to permit comparison of scores on different forms.

Apart from variations in difficulty level, the question of varying reliability of the forms may also be considered. In a study employing Forms D, E, and F, Wright and Laing (**6**) presented evidence that the forms vary in reliability. The nature of the experimental design of the present study does not permit calculation of interform reliability coefficients for all *Ss* because of the indeterminate effects of the counterbalancing on the correlation of any two forms. Interform coefficients based on the first and second forms administered to each *S* may, however, be computed validly. In the light of the paradigm shown in Table 1, it was possible to compute the product-moment coefficients shown in Table 5, each r being based on 14 cases. The corresponding standard errors of measurement are based on the average of the *SD*'s of the forms involved (using the values for the entire sample of 70 shown in Table 2).

The reliability coefficients and standard errors of measurement among Forms D, E, and F only may be compared with those reported by Wonderlic and Hovland (**5**), shown in Table 5 footnotes. Despite the restricted sample size, there is remarkably close agreement with the findings of Wonderlic and Hovland. Unfortunately no similar data involving Forms A and B have been published and no comparison with other studies is possible for the remaining coefficients of Table 5. However, to provide a check on the lowest coefficient, that of .63 for Forms B and D, these forms were administered to an independent similar sample of 28 cases. The resulting r was .65, which when combined with the previous 14 cases yields an r of .64. This

Table 3

Variance Table for Form Difference Data

Source of Variation	Sum of Squares	*df*	Variance Estimate
Rows: Individual differences*	9,626.1	69	139.5
Columns: Form differences**	1,030.1	4	257.5
Remainder: Error	3,525.9	276	12.78
Total	14,182.1	349	

* $F = 139.5/12.78 = 10.92$; $df = 69/276$; P less than .001.
** $F = 257.5/12.78 = 20.15$; $df = 4/276$; P less than .001.

would further indicate that the coefficients of Table 5 are stable.

It may be noted that the r's involving Forms A and B are lower than those involving D, E, and F alone, and the standard errors of measurement are higher. This would add further weight to the view presented above that Forms A and B appear to comprise a group having different properties from the group made up of D, E, and F. The small size of the samples on which the r's are based, however, makes interpretation tentative.

Conclusions

The Wonderlic forms vary in level of difficulty and cannot be regarded as equivalent and interchangeable. Separate tables of norms should be constructed for each form, or conversion formulas should be developed to permit comparison of scores. The forms appear also to vary in reliability.

Table 4

Form Differences and t-Test Results*

Forms	Difference	Probability
A and B	1.95***	less than .01
A and D	1.58***	less than .01
A and E	1.52**	less than .02
A and F	3.12†	less than .001
B and D	3.53†	less than .001
B and E	3.47†	less than .001
B and F	5.07†	less than .001
D and E	0.06	
D and F	1.54**	less than .02
E and F	1.60***	less than .01
Total	23.44	
Average	2.34†	less than .001

* 1.19 needed for significance at P .05 (276 df).
** 1.41 needed for significance at P .02.
*** 1.57 needed for significance at P .01.
† 2.02 needed for significance at P .001.

Table 5

Interform Reliability Coefficients and Standard Errors of Measurement

Forms	A–B	B–D	D–E	E–F	F–A
r	0.79	0.63*	0.86**	0.87***	0.80
$\sigma_{1_{xx}}$	3.04	3.99	2.11	2.00	2.50

* 0.64 when combined with independent similar sample of 28 cases (σ_r = 0.093).
** Compare with 0.85 and 0.88 reported by Wonderlic and Hovland (5).
*** Compare with 0.89 and 0.92 reported by Wonderlic and Hovland (5).

Received June 13, 1955.

References

1. Buros, O. K. (Ed.) *The third mental measurements yearbook.* New Brunswick, New Jersey: Rutgers Univer. Press, 1949. Pp. 347–350.
2. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
3. Melton, A. W. The methodology of experimental studies of human learning and retention. *Psychol. Bull.*, 1936, **33**, 305–394.
4. Wonderlic, E. F. *Wonderlic Personnel Test Manual.* Northfield, Illinois, 1945.
5. Wonderlic, E. F., & Hovland, C. I. The personnel test: a restandardized abridgment of the Otis S-A test for business and industrial use. *J. appl. Psychol.*, 1939, **23**, 685–702.
6. Wright, J. H., & Laing, D. M. The time factor in the administration of the Wonderlic Personnel Test. *J. appl. Psychol.*, 1943, **27**, 316–319.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Experimental Manipulation of the Halo Effect

Donald M. Johnson and Robert N. Vidulich

Michigan State University

The halo effect was briefly noted in 1907 by Wells (**6**) and christened in this Journal in 1920 by Thorndike (**5**). Since then nearly all discussions of ratings have listed this effect as one of the errors to be avoided. Yet when the evidence was reviewed in 1945 (**2**) and again in 1955 (**3**), no convincing proof of the existence of halo could be found. The present report is offered as the first published verification of this 50-year-old concept.

There is no doubt about the facts to which this luminous phrase is applied. When anyone is asked to rate several individuals on several traits, he will rate some individuals high on most traits and some low on most traits. If the ratings of the individuals on any two traits are correlated, the correlation will be positive and often quite large.

There is some doubt about the interpretation of the facts. Conventionally, the correlation between trait ratings is considered an error of judgment, at least in part. The rater has a generally favorable or unfavorable attitude toward each individual that influences his ratings of the individual on each trait. The favorable attitude is the one, of course, that gave halo its name. After the rater has cast a halo around his subject, he is so dazzled by its radiance that he cannot differentiate the subject's separate qualities. If the rater judged the individuals on each trait separately and analytically, as requested, the traits would not be correlated.

The criticism of this interpretation is that the rater may be right. Perhaps the correlation between the traits is an objective fact rather than an error of judgment (**2**). Or at least it may be the information available to the rater rather than his judgment that is in error. The correlation between trait ratings may be due to an objective correlation, to correlation in the rater's information, or to the rater's inability to judge the traits separately, i.e., a halo effect. Statistical analysis of trait ratings will not by itself locate the source of the correlation.

The common recommendation for avoiding the halo effect gives a clue to the proper method of investigation. Several writers, e.g., Symonds (**4**), have recommended rating all individuals on one trait at a time, rather than one individual at a time on all traits. Presumably this will lessen confusion between traits and thus reduce the halo effect. Such a reduction in the halo effect would support the hypothesis that the effect resides in the judging process rather than in the objective facts.

One reason why there is little evidence on halo is the number of correlation coefficients involved. If someone rates a number of individuals on five traits, 10 intertrait correlations have to be computed. If 10 raters are used, 100 correlations are required. If the ratings are made under two conditions, the experiment involves 200 correlations.

Fortunately, the more efficient analysis of variance has recently been applied to the entire treatment of ratings by Guilford (**1**). His elegant rationale for errors of rating simplifies the calculations and clarifies the interpretation of the results. He distinguishes between two components of the halo effect, that common to different raters and that varying from rater to rater. If the ratees, or individuals being rated, are not all rated equally, some getting higher ratings than others, the analysis discloses a significant variance between ratees. This between-ratee variance is the more "objective" component of the halo effect in the sense that it is common to the different raters, who may all have been exposed to the same information. If some raters rate some ratees higher than others, the analysis discloses a significant interaction variance, which Guilford calls the relative halo effect. It is relative to the raters, traits, and ratees at hand and is estimated in relation to

Table 1

Mean Ratings of Five Prominent Individuals on Five Traits by 18 College Students Under Conditions that Maximize the Halo Effect

Ratee	Intelligence	Personal appearance	Kindliness	Courage	Usefulness	Means
Queen Elizabeth	8.50	7.94	7.67	7.33	6.17	7.52
Senator McCarthy	8.06	6.00	3.67	8.22	5.28	6.25
Sir Winston Churchill	9.17	6.17	8.00	8.44	8.89	8.13
Mrs. Eleanor Roosevelt	8.17	6.00	7.72	7.22	6.28	7.08
Pope Pius XII	9.17	7.50	8.56	8.56	8.50	8.46
Means	8.61	6.72	7.12	7.96	7.02	7.49

deviations from the grand mean of all ratings by all raters.

These two components of the halo effect calculated by analysis of variance are, of course, open to the same questions as the halo effect calculated by correlations between trait ratings. To prove that these halo effects, or some part of them, are due to an error of judgment rather than to objective facts, we may experimentally manipulate the conditions of judgment, while holding information about the facts constant, and compare the variances obtained under the different conditions.

Procedure

Two groups of raters were used, one working under conditions designed to maximize the halo effect, the other working under conditions designed to minimize the halo effect. The ratees were five well-known living persons. The traits were five varied but familiar characteristics. Ratings were made on a scale from 1 to 10, with 10 being high. Since five ratings were made and the classes met two days a week, the experiment covered two and a half weeks. The classes were treated the same except as described below.

The maximization group was given on the first day a list of traits and the name of one ratee, as follows:

QUEEN ELIZABETH

Intelligence ——
Personal appearance ——
Kindliness ——
Courage ——
Usefulness ——

They were told to rate this individual in respect to each trait. Examples were given. At the next class meeting they were given another ratee and the same list of traits. And so on for five ratees.

The minimization group was given on the first day a list of ratees and one trait, as follows:

INTELLIGENCE

Queen Elizabeth ——
Senator Joseph McCarthy ——
Sir Winston Churchill ——
Mrs. Eleanor Roosevelt ——
Pope Pius XII ——

They were asked to rate each individual in respect to this trait. Examples were given. At the next class meeting they were given another trait and the same list of ratees. And so on for five traits.

All raters in each group made their judgments in the same order. These orders are shown in the lists printed above.

The hypothesis was that rating one individual on all traits at once would increase the halo effects, and rating one trait at a time would reduce them. That is, the minimization group followed the procedure commonly recommended for reducing the halo effect.

The raters were students of two classes in Effective Study at Michigan State University. The maximization group began with 24, the minimization group with 23. Students who were absent from any of the five experimental days were eliminated, so both groups were reduced to 18 raters: 14 men and 4 women in the maximization group, 13 men and 5 women in the minimization group. In respect to intelligence (ACE), college class, age, a rough measure of family income, and religion, the groups were closely comparable. No explanation of the purpose of the experiment was given. The students were accustomed to simple classroom projects of this kind.

Results

In general, the ratings shown in Tables 1 and 2 are about as expected. Sir Winston Churchill was rated very high on intelligence, Senator McCarthy very low on kindliness. The largest standard deviation was for Senator McCarthy in respect to usefulness. The grand means were approximately the same under both conditions, as were the total variances.

Table 2

Mean Ratings of Five Prominent Individuals on Five Traits by 18 College Students Under Conditions that Minimize the Halo Effect

Ratee	Intelligence	Personal appearance	Kindliness	Courage	Usefulness	Means
Queen Elizabeth	6.67	8.94	7.44	6.72	6.28	7.21
Senator McCarthy	6.78	6.89	4.00	7.83	6.06	6.31
Sir Winston Churchill	9.06	6.44	6.56	8.33	8.72	7.82
Mrs. Eleanor Roosevelt	7.83	6.28	7.78	6.78	6.33	7.00
Pope Pius XII	8.06	7.72	8.17	8.33	7.83	8.02
Means	7.68	7.25	6.79	7.60	7.04	7.27

The important results for our purposes are the two analyses of variance. Table 3 gives the results for the maximization group, Table 4 for the minimization group. Variance due to raters is significant beyond the .01 level under both conditions. The largest variance is due to ratees or individuals being rated. This is the component of the halo effect that is common to all raters. It is considerably smaller under the minimization condition.

The relative halo effect is in the variance due to interaction between raters and ratees (R × I), shown in Part I of each table. This variance is small under both conditions, but under maximization conditions it is significant beyond the .01 level of confidence. Under minimization conditions it does not reach the .05 level. This verifies the long-standing assumption that part of the halo effect is a phenomenon of judgment and also shows that the common recommendation about removing halo is effective. When the traits were rated one at a time, two or three days apart, the relative halo effect was not significant.

As a check on this difference between conditions, the data summarized in Part I of each table were put into an 18 × 5 × 2 design to

Table 3

Analysis of Variance of Ratings of Five Prominent Individuals on Five Traits by 18 College Students Under Conditions that Maximize the Halo Effect

I. Ignoring Differences Between Traits

Source	Sum of Squares	Degrees of Freedom	Variance Estimate	*F*	*P*
Between raters (R)	251.78	17	14.81	5.16	.01
Between ratees (I)	276.16	4	69.04	24.06	.01
Interaction (R×I)	365.68	68	5.38	1.87	.01
Within sets	1032.80	360	2.87		
Total	1926.42	449			

II. Ignoring Differences Between Individuals

Source	Sum of Squares	Degrees of Freedom	Variance Estimate	*F*	*P*
Between raters (R)	251.78	17	14.81	4.05	.01
Between traits (T)	217.54	4	54.39	14.86	.01
Interaction (R×T)	138.70	68	2.03	.56	n.s.
Within sets	1318.40	360	3.66		
Total	1926.42	449			

Table 4

Analysis of Variance of Ratings of Five Prominent Individuals on Five Traits by 18 College Students Under Conditions that Minimize the Halo Effect

I. Ignoring Differences Between Traits

Source	Sum of Squares	Degrees of Freedom	Variance Estimate	*F*	*P*
Between raters (R)	243.14	17	14.30	4.04	.01
Between ratees (I)	167.99	4	42.00	11.86	.01
Interaction (R×I)	268.65	68	3.95	1.12	n.s.
Within sets	1273.60	360	3.54		
Total	1953.38	449			

II. Ignoring Differences Between Individuals

Source	Sum of Squares	Degrees of Freedom	Variance Estimate	*F*	*P*
Between raters (R)	243.14	17	14.30	3.96	.01
Between traits (T)	50.19	4	12.55	3.48	.01
Interaction (R×T)	358.85	68	5.28	1.47	.05
Within sets	1301.20	360	3.61		
Total	1953.38	449			

permit calculation of the triple interaction, raters × ratees × conditions. This interaction was significant beyond the .01 level. The interaction between ratees and conditions was also significant beyond the .01 level.

The interaction between raters and traits (R × T) shown in Part II of both tables, has not attracted as much attention as the halo interaction. This interaction indicates the tendency of some raters to give high ratings on some traits and low ratings on other traits. As Guilford points out, this interaction could include a contrast effect referred to the self. That is, someone might rate everyone low on courage because he considers himself unusually courageous. Or, in reverse, someone might rate everyone high on kindliness because he is a kindly person himself. Whatever is included in this rater-trait interaction, it is affected by our conditions in just the opposite way from the halo interaction. It is significant at the .05 level under the conditions designed to minimize halo and is hardly present at all under the maximization condition.

As a check on this difference, the data summarized in Part II of each table were put into an 18 × 5 × 2 design to permit calculation of the triple interaction, raters × traits × conditions. This interaction did not reach the .05 level of significance.

Though this shift in R × T interaction between conditions is not significant and was not expected in advance, it seems to fit in with the other results. When the rater judges one individual at a time on all traits, he manages to treat the traits in isolation but he gets personally involved with the individuals. Similarly, when the rater judges all individuals on one trait at a time, he manages to treat the individuals separately but he gets personally involved with the traits. In reducing the relative halo or rater-ratee interaction we have increased the rater-trait interaction, in this sample at least.

It is possible to state a generalization that encompasses both interactions. The over-all interaction is between raters and what they do on separate days. The rater can attend to, and judge separately, five items presented simultaneously, but when these items are presented a few days apart as aspects of a larger task, he cannot treat them separately.

The third interaction, between traits and

ratees, computed by ignoring differences between raters, was significant beyond the .01 level under both conditions. It seems to be of no particular importance to the present discussion.

Limitations

Though we have used the term "maximization," we have no experience or data to indicate that relative halo variance in Table 3 is really maximal. Our selections of ratees, traits, and rating conditions were based on current hypotheses about halo, with no data to guide these selections. It is possible that halo would be larger if the ratings were made at intervals of one day, for example, or one hour.

The same considerations apply to the small interaction variance between raters and traits, and the small difference in this variance between conditions. It is doubtful that much error of any practical importance enters these ratings because of this interaction. Under other conditions, of course, this interaction may be larger and may shift more. The effect of the judging conditions should therefore be considered when anyone uses a rater's ratings to get information about the rater.

Summary

The halo effect is supposed to be due to an error of judgment but no evidence has been published to prove that correlations between trait ratings are due to errors of judgment rather than objective correlations. The correct method of investigation is experimental variation of the conditions of judgment and comparison of halo effects manifested under the different conditions.

One group rated five well-known individuals on five traits under conditions designed to maximize the halo effect. They rated one individual each experimental day on all traits. The other group rated the same individuals on the same traits under conditions designed to minimize the halo effect. They rated all individuals on one trait each experimental day.

The relative halo effect was calculated, following Guilford, as the variance due to interaction between rater and ratee. Under maximization conditions this interaction variance was significant beyond the .01 level. Under minimization conditions this interaction variance was not significant at all. The between-ratee variance also was smaller under the minimization conditions. These results prove that halo is in part a phenomenon of judgment and that the common recommendation for reducing it by changing the conditions of judging is effective.

When halo variance was reduced by this change in the judging task, variance due to interaction between raters and traits was increased, but this difference was not significant.

In general, the rater attends to and differentiates the five items spread out in front of him whether they are names of individuals or traits. He interacts personally with these same items when they are presented at intervals of two or three days as aspects of a larger task.

Received May 5, 1955.

References

1. Guilford, J. P. *Psychometric methods.* (2nd Ed.) New York: McGraw-Hill, 1954.
2. Johnson, D. M. A systematic treatment of judgment. *Psychol. Bull.,* 1945, **41,** 193–224.
3. Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.
4. Symonds, P. M. Notes on rating. *J. appl. Psychol.,* 1925, **9,** 188–195.
5. Thorndike, E. L. A constant error in psychological ratings. *J. appl. Psychol.,* 1920, **4,** 25–29.
6. Wells, F. L. A statistical study of literary merit. *Arch. Psychol.,* 1907, **1,** No. 7.

The Journal of Applied Psychology
Vol. 40, No. 2, 1956

Comparison of Two Visual Display Presentations [1]

A. V. Churchill

Defense Research Medical Laboratories, Toronto, Ontario

The applicability of the results of numerous studies of dial legibility is dependent on the assumption that the presentation of slides on a screen is equivalent to the presentation of the actual dials. The present experiment was conducted in order to establish the comparability of the two modes of presentation.

Method

Apparatus. Twelve 100 × 2, black on white dials, three inches in diameter, designated as types "A" and "B" in a previous study (**1**), were arranged in two simulated six-dial panels, with a spring-loaded pointer mounted on each dial. The panels were presented in a vertical frame, at a 30-inch viewing distance. The panels were viewed through an aperture, fitted with a shutter which the subject (*S*) operated by means of two microswitches.

Slides made from photographs of the panels were projected onto a vertical screen which presented 6-inch diameter dials at a 60-inch viewing distance. As with the panels, the slide exposures were controlled by *S*.

Table 1

Comparison of Slide and Panel Presentations

	Slides	Panels
Mean reading time per dial	1.46 sec.	1.42 sec.
% of readings in error	3.92%	3.85%
Mean error per reading (in scale units)	+0.056	+0.048
Errors of more than 10 units*	20	13

* Thirty-three errors of a magnitude greater than 10 were treated as errors of 10, i.e., 91 called 51, 91 called 9.

The ambient illumination was 20 foot-candles, the illumination at the display was 70 foot-candles during presentation.

Subjects. The *S*s were 50 University undergraduates, with normal vision, who were paid for their participation.

Procedure. Forty-eight pointer positions were presented to each *S*, twice on the panels and twice on the slides. In both test situations *S* was seated in front of the display and held a microswitch in either hand. At the "ready" signal *S* pressed the "on" switch, reported the six dial settings as quickly and as accurately as possible, and then pressed the "off" switch. Twenty seconds was allowed between the reporting of the readings on one presentation and the "ready" signal for the following presentation. Panels preceded or followed slides for alternate *S*s, and a different random order of settings, on both slides and panels, was presented to each *S*.

Results and Conclusions

Reading time and error data were recorded. The results are presented in Table 1.

As will be seen from the table, there is a slight, but insignificant advantage in favor of the panel presentation, the time advantage for panels showing a t of .75.

Analysis of the data which were gathered in the present study lends justification to the procedure of applying slide-projection data directly to the actual displays which the slides are intended to represent.

Received May 9, 1955.

Reference

1. Churchill, A. V., & Allan, D. G. Experimental dial design. March, 1955. *Defense Res. Bd., Canada, DRML Rep.* No. 164–1.

[1] Defense Research Medical Laboratories Report No. 164–2, Project No. D77–94–20–27 (H.R. No. 84).

"Keep up" with reviews of the best, current books in applied psychology—

Read *Contemporary Psychology*

The *New* APA Journal of Book Reviews

Let *CP* provide you with the current, monthly news of books in the broad field of psychology, as well as in the various phases of applied psychology—now that the *Journal of Applied Psychology* is no longer publishing book reviews.

—Every month *CP* publishes one long, critical, essay-type review of a particularly important psychological volume.

—Every month *CP* presents many well-written, evaluative reviews of the foremost new volumes.

—Every month *CP* includes valuable information on many phases of educational and industrial psychological films . . . the films themselves . . . film research . . . books about films.

—Every month in "*CP* Speaks" the editor provides news and comment on psychology's wide world of books.

Read CP *for the News in Books*

$4.00—to APA members; and participants in Student Journal Groups

$8.00**—to all others (add **$.50 for foreign postage, please)

American Psychological Association

1333 Sixteenth Street, N. W. **Washington 6, D. C.**

Features systematic examination of union-management relations as a psychological problem . . .

THE PSYCHOLOGY OF INDUSTRIAL CONFLICT

By Ross Stagner, University of Illinois. This remarkable book presents an entirely new approach to the understanding of industrial controversy. It offers an analysis of why workers join unions, why executives follow certain policies in labor relations, leadership in unions and management, tactics, strike phenomena, and labor-management cooperation—in terms of the behavior of the individual.

Although not a guide to labor-relations, the book does provide a new way of looking at issues that can be of considerable help in formulating policy. It portrays the competition between unions and management as a complex set of interactions occurring at the individual, group, and institutional levels. It brings together all the relevant material on the psychological factors involved, formulates general principles, and demonstrates the usefulness of these principles with actual examples. *1956. 550 pages. $8.00.*

PSYCHOSOMATICS

By Max Hamilton, University of Leeds School of Medicine. Delves into the wide variety of illnesses that center in the respiratory and gastrointestinal systems and, in addition, discusses cardiovascular conditions, disorders of the skin, endocrine and metabolism disturbances, and rheumatic ailments. Psychosomatic theory is discussed at length. *1955. 225 pages. $4.25.*

Send for examination copies.

JOHN WILEY & SONS, Inc. 440-4th Ave., New York 16, N. Y.

Outstanding

McGRAW-HILL

Books

INTRODUCTION TO PSYCHOLOGY

By CLIFFORD T. MORGAN, The Johns Hopkins University. 694 pages, $6.00

This new and different text presents the principles and applications of modern psychology in readable and interesting form. It is intended for introductory college courses. Comprehensive in coverage, it offers a broad perspective of psychology, including both its physiological and social aspects. It is thoroughly up to date and abreast of developments in the different areas of psychology. Includes an abundance of excellent illustrations.

PSYCHOLOGY IN MANAGEMENT

By MASON HAIRE, University of California. 220 pages, $4.75 (*text edition available*).

The emphasis in this new book is on the psychological principles underlying industrial problems. An industrial psychologist discusses the nature of people, and how their needs express themselves and press for satisfaction. It presents a coherent, integrated picture of psychological problems in application to managerial problems in industry.

COUNSELING PSYCHOLOGY

By MILTON E. HAHN and MALCOLM S. MACLEAN, University of California, Los Angeles. Second Edition. 312 pages, $4.75

A carefully revised, improved, and fully up to date edition of the authors' GENERAL CLINICAL COUNSELING. Covers both the knowledge and the skills necessary for the practice of counseling psychology. A combination of the "why" and "how" of counseling psychology . . . presenting an insight into trait and factor, personality, learning, and perception theories, into group dynamics, and statistical orientations as well . . . all requirements for effective practice. Attention is given to principles and the ethical code of the American Psychological Association.

PERSONNEL AND INDUSTRIAL PSYCHOLOGY

By EDWIN E. GHISELLI and CLARENCE W. BROWN, University of California, Berkeley. *McGraw-Hill Series in Psychology.* Second Edition. 506 pages, $6.00

The major purpose of this book is to provide a comprehensive treatment of personnel and industrial psychology, and information relative to principles and procedures that have a bearing on the more effective utilization of manpower. This second edition is thoroughly revised and contains new chapters on selection and classification of workers and on social factors in industry. Principles rather than practices are emphasized throughout, and the importance of empirical verifications is stressed.

Send for copies on approval

McGraw-Hill
BOOK COMPANY, INC.
330 West 42nd Street New York 36, N. Y.